The conformational space and dynamics of the partially disordered transcription factor engrailed-2 explored with magnetic resonance

Shahid Nawaz Khan

École Doctorale 388: Chimie Physique et Chimie Analytique

Exploration par résonance magnétique de l'espace conformationnel et de la dynamique du facteur de transcription partiellement désordonné Engrailed-2

THESE DE DOCTORAT DE L'UNIVERSITE PIERRE ET MARIE CURIE

Spécialité:

RMN BioMoléculaire

Shahid Nawaz Khan

Soutenue le 12 mars 2015

Devant un jury composé de:

Président du jury	Prof. Germain Trugnan
Directeur de thèse	Dr. Fabien Ferrage
Co-directeur de thèse	Prof. Olivier Lequin
Rapporteurs	Assoc Prof. Isabella C. Felli
	Dr. Carine Tisne
Examinateurs	Dr. Martin Blackledge

UPMC
SORBONNE UNIVERSITÉS

École Doctorale 388: Chimie Physique et Chimie Analytique

The Conformational Space and Dynamics of the Partially Disordered Transcription Factor Engrailed-2 Explored with Magnetic Resonance

THESE DE DOCTORAT DE L'UNIVERSITE PIERRE ET MARIE CURIE

Spécialité:

RMN BioMoléculaire

Shahid Nawaz Khan

Soutenue le 12 mars 2015

Devant un jury composé de:

Président du jury	Prof. Germain Trugnan
Directeur de thèse	Dr. Fabien Ferrage
Co-directeur de thèse	Prof. Olivier Lequin
Rapporteurs	Assoc Prof. Isabella C. Felli
	Dr. Carine Tisne
Examinateurs	Dr. Martin Blackledge

The Conformational Space and Dynamics of the Partially Disordered Transcription Factor Engrailed-2 Explored by Magnetic Resonance

It Is Best To Learn As We Go, Not Go As We Have Learned.
-- Leslie Jeanne Sahle --

Acknowledgment

Abstract (English)

Intrinsically Disordered Proteins (IDPs), which lack a stable rigid structure constitute a large and functionally important class of proteins. Nuclear Magnetic Resonance (NMR) is a well-established technique to characterize the structural and dynamical features of IDPs at atomic resolution. The broad conformational space of IDPs makes them challenging targets for structural biology to define their precise structural features and motions, the physical and chemical properties that underlie their biological functions. The present thesis establishes biophysical investigation of the disordered region of the transcription factor Engrailed-2 (13.5 kDa) primarily by NMR. After describing the protocol of expression and purification of the isotopically labeled protein, we present a novel approach to characterize the pico – nano second motions in IDPs using nuclear spin relaxation data at multiple fields. Paramagnetic Relaxation Enhancements (PREs) are used to identify transient long-range interactions between the disordered region and the folded homeodomain of Engrailed-2. Binding to DNA was studied by fluorescence anisotropy and highlights the role of the disordered region in the DNA binding. We used Electron Paramagnetic Resonance (EPR) to probe the potential interaction between the hydrophobic cluster (hexapeptide) in the disordered region and the homeodomain. The one-bond ^{1}H-^{15}N, C_α-H_α and C_α-C′ residual dipolar couplings (RDCs) measured for Engrailed-2 provide important constraints for the refinement of the conformational space of Engrailed_2. All these approaches provide valuable insights in understanding the structural, dynamical and functional properties of this IDP.

Key Words: Intrinsically Disordered Protein, Nuclear Magnetic Resonance, Homeoproteins, Protein dynamics, Paramagnetic Relaxation Enhancement, Protein-DNA interaction, Residual Dipolar Couplings.

Abstract (French)

Les protéines intrinsèquement désordonnées (IDP), dépourvues d'une structure rigide et stable, constituent une classe de protéines diverses et fonctionnellement importantes. La résonance magnétique nucléaire (RMN) est une technique spectroscopique bien établie pour caractériser les propriétés conformationnelles et dynamiques des IDP avec une résolution atomique. L'espace conformationnel, en général large et varié, des IPD en fait une cible difficile pour la biologie structurale dont le but est de déterminer avec précision et exactitude les propriétés structurales, dynamique et physico-chimiques qui sous-tendent la fonction des macromolécules biologiques. Ce manuscrit présente une étude biophysique détaillée de la région intrinsèquement désordonnée (IDR) du facteur de transcription Engrailed-2, avant tout par RMN. Après une présentation de cette homéoprotéine, nous décrivons les protocoles d'expression et de purification de cette protéine isotopiquement marquée. Nous introduisons ensuite une nouvelle approche pour la caractérisation des mouvements pico- et nanoseconde des protéines intrinsèquement désordonnées à partir de données de relaxation des spins nucléaires enregistrées à plusieurs champs magnétiques. Les effets de relaxation paramagnétique (PRE) ont été utilisés pour identifier des interactions transitoires entre la région désordonnée et l'homéodomaine d'Engrailed-2. L'interaction d'Engrailed-2 avec l'ADN a été étudiée en détail en utilisant l'anisotropie de fluorescence sur une série de constructions de la protéine, afin de mettre en lumière le rôle de la partie désordonnée dans l'interaction avec l'ADN. Nous avons également employé la résonance paramagnétique électronique pour tenter de détecter une interaction potentielle entre le noyau hydrophobe de l'hexapeptide dans la région désordonnée et l'homéodomaine. Les couplages dipolaires résiduels (RDC) dans les paires ^{1}H-^{15}N, C_α-H_α et C_α-C' ont également été mesurés sur des échantillons d'Engrailed en milieu anisotrope. Ces données seront essentielles pour reconstituer l'espace conformationnel d'Engrailed 2. L'ensemble des approches présentées a permis de constituer un socle solide de connaissances qui permettent de mieux comprendre les propriétés conformationnelles, dynamiques et fonctionnelles de l'IDR d'Engrailed-2.

Mots-clés: Protéines intrinsèquement désordonnées, Résonance magnétique nucléaire, Homéoprotéines, Dynamique des protéines, Relaxation paramagnétique, Interaction Protéine-ADN, Couplages dipolaires résiduels

Contents

Summary (French)

Exploration par résonance magnétique de l'espace conformationnel et de la dynamique du facteur de transcription partiellement désordonné Engrailed-2

Les protéines sont des constituants fondamentaux de toutes les cellules vivantes, qui sont requis pour la structure, la fonction et la régulation des cellules des organismes, des tissus, et des organes. Parmi les exemples de protéines on peut citer les enzymes, les hormones, et les anticorps qui jouent un rôle essentiel dans le fonctionnement correct d'un organisme. La diversité structurale des protéines leur permet d'accomplir une grande variété de fonctions.

Séquence en acides aminés → Structure Tridimensionnelle → Fonction protéique

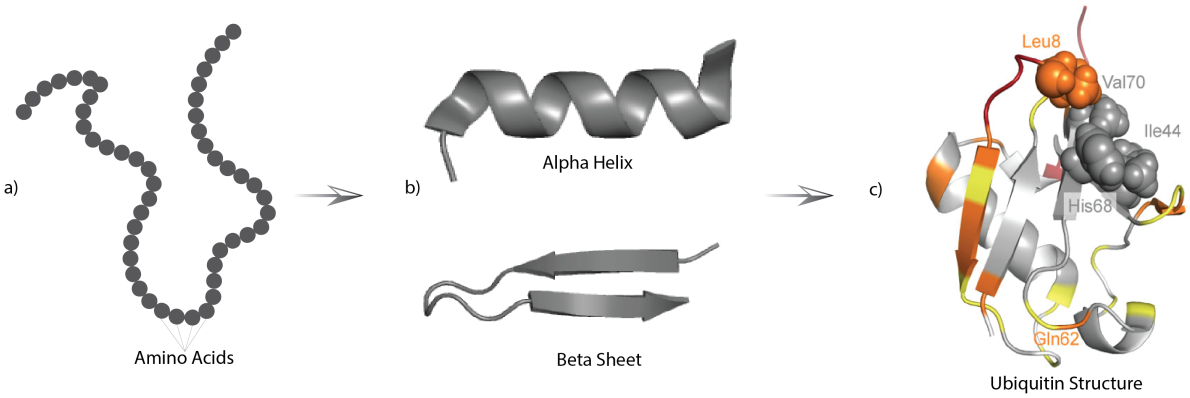


Figure 1: (a) a) Les protéines sont constituées de longues chaînes d'acides aminés. (b) Ces chaînes polymériques monodimensionnelles se replient en général en adoptant une variété étonnante motifs de structuraux (structure secondaire). (c) Ces éléments sont associés pour former des structures 3D dynamiques, qui permettent aux protéines d'assurer leur grande diversité de fonctions.

Le paradigme structure/fonction des protéines a été érigé en dogme central de la biologie structurale et de la biochimie des protéines pendant plusieurs décennies, énonçant que la fonction d'une protéine dépend strictement de l'acquisition préliminaire d'une structure 3D. La structure d'une protéine est stabilisée par une

variété d'interactions intramoléculaires non covalentes comme les liaisons hydrogène, les interactions électrostatiques et les forces de van der Waals. Ces interactions peuvent être affaiblies et la structure protéique peut ainsi être déstabilisée par des agents dénaturants conduisant au dépliement de la protéine.[1,2] Selon les conditions dénaturantes et les propriétés intrinsèques de la protéine, plusieurs états non repliés peuvent être observés, pouvant aller d'états partiellement dépliés retenant une majorité de la structure native à des états complètement désordonnés proches d'une chaîne statistique ("random coil") ne possédant aucune structure secondaire en hélice α ou feuillet β.[3,4] Il y a maintenant plus de 30 ans, la première entorse au dogme de la nécessité d'une structure 3D pour accomplir une fonction, est apparue avec la découverte de segments de protéines n'ayant pas de structure 3D stable mais étant pourtant parfaitement fonctionnelles. Ces segments sont appelés désormais protéines intrinsèquement désordonnées (IDP pour "Intrinsically Disordered Proteins") ou régions intrinsèquement désordonnées (IDR pour "Intrinsically Disordered Regions").[5]

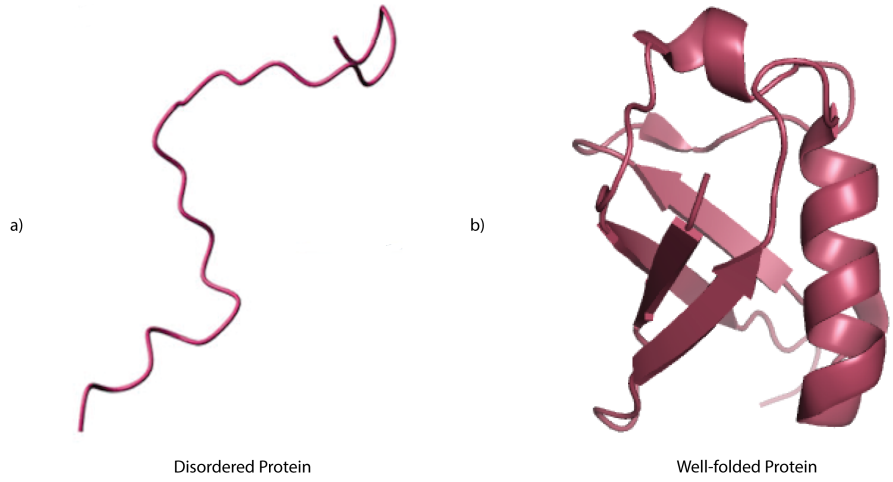


Figure 2: (a) Protéine sans structure 3D définie ou bien ordonnée appelée IDP. (b) Protéine ayant une structure 3D bien définie (Ubiquitine; code pdb 1d3z).

L'existence ubiquitaire des IDP/IDR, n'ayant pas de structure tridimensionnelle bien définie mais jouant néanmoins des rôles cruciaux dans un grand nombre de processus cellulaires, a non seulement défié le paradigme structure–fonction mais a

aussi ouvert la voie vers une nouvelle classe d'objets d'étude intrigants pour la protéomique moderne.[6,7] Les IDP et les IDR sont impliqués dans une grande variété de processus biologiques, allant de la régulation et la signalisation au contrôle cellulaire et qui complètent les fonctions des protéines ordonnées. Par exemple, les IDP p21 et p27 se lient aux kinases dépendantes des cyclines (cdk) et fonctionnent ainsi comme des régulateurs du cycle des cellules de mammifères via des interactions directes avec des complexes Cdk/cycline.[8] Le facteur de transcription Myc qui se lie à l'ADN pour réguler la division de nombreux types cellulaires est intrinsèquement désordonné à l'état isolé et se replie lors de la liaison à l'ADN.[9] L'α-synucléine, une IDP dans l'état non lié, est potentiellement impliquée dans le réarrangement synaptique et agit également comme un chaperon moléculaire dans l'activité synaptique.[10] De nombreuses maladies humaines, incluant le cancer, les maladies amyloïdes, le diabète, les maladies cardiovasculaires et neurodégénératives[11] sont associées à des défauts de reconnaissance et de signalisation de certaines IDP. Le développement de stratégies nouvelles de conception de médicaments ciblant les IDP reflète leur implication dans ces maladies.[12]

Structure des IDR/IDP

La caractérisation des IDR/IDP par des méthodes computationelles et expérimentales (RMN, SAXS, smFRET etc.) démontre la grande diversité des fluctuations conformationnelles pour classer les IDR/IDP en de nombreux types structuraux. La nature hautement dynamique des IDP et leur capacité à fluctuer rapidement entre différents états (random coil, structure secondaire, structure globulaire) font de ces fluctuations structurales un espace conformationnel continu. En effet l'ensemble hétérogène de conformations allant d'états random coil à des structures compactes permet de définir les IDP comme un ensemble continu de structures en interconversion rapide.[13]

Le modèle à 4 états énonce que les protéines peuvent exister à l'état random coil, pré-molten globule, molten globule et un état complètement replié. Les transitions des protéines repliées pour atteindre des états fonctionnels renforcent le besoin d'une

adaptation structurale pour accomplir une fonction spécifique.[14] Les protéines dans un état replié ou ordonné sont des objets compacts possédant une structure secondaire et tertiaire tandis que les états “molten globule“ et “pré-molten globule“ ont des structures moins compactes mais possèdant toujours une structure secondaire résiduelle. En revanche, l’état “random coil“ ne possède pas ou peu de structure secondaire.[15] Il est clair désormais d’un point de vue expérimental que les IDR/IDP peuvent adopter ces différents états à l’exclusion de l’état replié, ce qui suggère que ces protéines peuvent être considérées comme adoptant un continuum de différents états désordonnés.

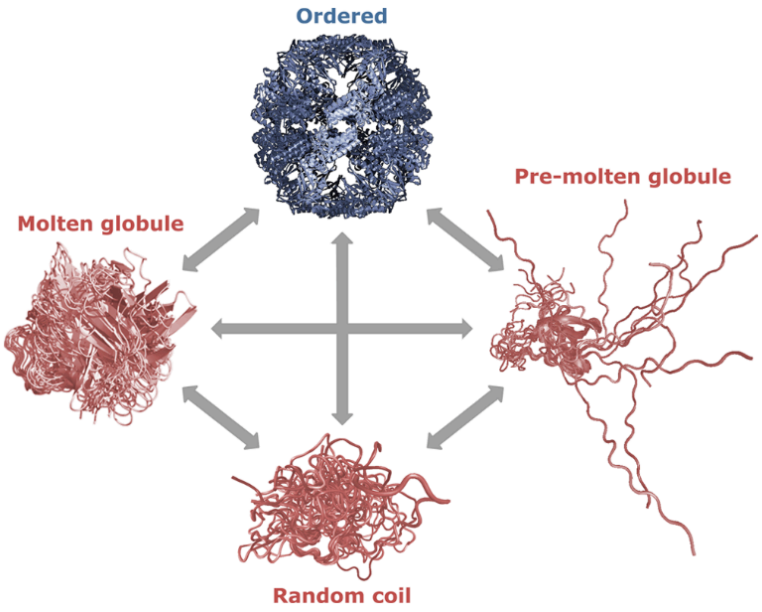


Figure 3: Modèle à 4 états de la structure des protéines. Les protéines peuvent adopter des états conformationnels spécifiques (état ordonné, molten globule, prémolten globule et random coil) pour accomplir leurs différentes fonctions). Reproduit depuis réf. 16.[16]

La plasticité structurale et le grand espace conformationnel des IDP/IDPR leur permet d’interagir avec de nombreux partenaires de liaison et d’être impliquées dans des réseaux de régulation transitoire.[17] La possibilité pour ces protéines d’acquérir des éléments de structure secondaire et tertiaire lors de la liaison aux partenaires biologiques est encore mal comprise en bien des aspects structuraux et dynamiques.[18] L’ensemble hétérogène de conformations allant de conformères étendus à des structures compactes conduit à les considérer comme un ensemble

continu de structures en interconversion rapide. De manière intéressante, une étude combinée de l'IDP ostéopontine par Résonance Paramagnétique Electronique (RPE) et par Résonance Magnétique Nucléaire (RMN) montre que certains éléments structuraux sont stabilisés par des interactions longue-distance similaires à celles que l'on trouve dans les protéines repliées. Cette étude combinée de l'ostéopontine démontre que la forme apo est peuplée à la fois par des états étendus et des états se repliant de manière coopérative.[13] La caractérisation de l'espace conformationnel et la description quantitative de la dynamique des IDP au niveau atomique visent à mieux comprendre leur fonction biologique.

La cristallographie des rayons X, qui est la technique majeure pour déterminer la structure des protéines, n'est pas adaptée pour caractériser structuralement les IDP et IDR. L'absence de densité électronique pour des atomes de protéines ne diffractant pas les rayons X de façon cohérente est généralement interprétée comme un indice de désordre, les IDR se caractérisent donc par l'absence d'information en cristallographie. Par leur nature, les IDPs sont difficilement cristallisables et, quand bien même un cristal serait obtenu, il ne représenterait au mieux qu'une unique structure de l'ensemble conformationnel. La diffraction des rayons X aux petits angles (SAXS) peut distinguer des protéines repliées et désordonnées car l'intensité et le profil des rayons X diffusés par l'échantillon à l'état liquide permet de déterminer la forme et la taille des protéines.[19] Il s'agit d'une contrainte importante pour caractériser l'espace conformationnel. La technique de transfert résonnant d'énergie de fluorescence sur molécule unique (smFRET) peut fournir une information sur la structure et sur des changements conformationnels dynamiques au niveau d'une molécule unique de protéine.[20] La microscopie de force atomique (AFM) est également utile pour observer la structure dynamique des protéines et peut sonder l'hétérogénéité conformationnelle en temps réel.[21]

L'étendue du désordre n'est pas une spécificité de la protéine ; même un acide aminé dans une protéine globalement statique fluctue à des degrés divers d'hétérogénéité structurale. L'hétérogénéité structurale et dynamique des IDP fait de la RMN une méthode unique pour explorer l'espace conformationnel et la cinétique d'exploration à une résolution atomique.[22] La spectroscopie RMN est une technique puissante de

biologie structurale. La RMN est hautement adaptée pour l'étude des IDP et IDR de par sa capacité unique à fournir une information avec une résolution atomique. La richesse en information contenue dans les déplacements chimiques, les vitesses de relaxation, les couplages scalaires et dipolaires et les vitesses de relaxation paramagnétique (PRE) fait de la RMN la technique de choix pour analyser les propriétés conformationnelles et dynamiques des IDP.

Résonance Magnétique Nucléaire

La RMN est une méthode extrêmement performante pour explorer la structure et la dynamique des biomolécules, en particulier en solution dans des conditions proches des conditions physiologiques. Les échantillons de biomolécules enrichis en isotopes actifs magnétiquement comme les noyaux ^{1}H, ^{13}C, ^{15}N or ^{31}P sont placés dans un champ magnétique intense et peuvent être explorés avec une résolution atomique.

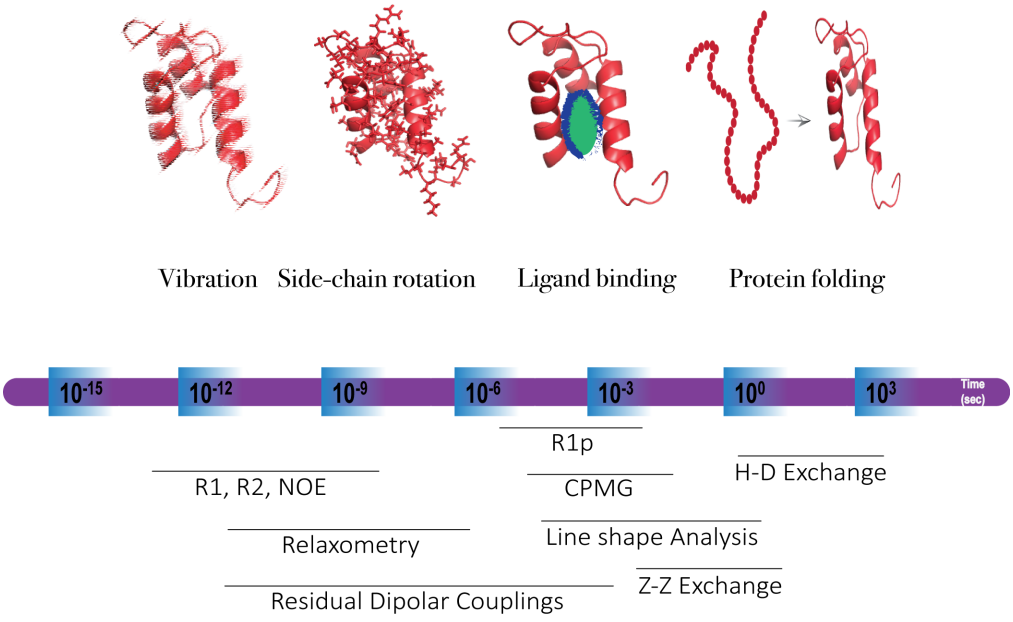


Figure 4: Les différentes échelles de mouvements dans les protéines qui peuvent être caractérisées efficacement par RMN. La détection des mouvements sur des échelles de temps de la picoseconde à plusieurs heures fait de la RMN un outil unique pour comprendre les fonctions biologiques à l'échelle atomique.

La puissance de la RMN réside dans sa possibilité de caractériser l'environnement chimique des noyaux, ce qui en fait une technique largement répandue dans le domaine de la chimie moléculaire, des matériaux, des sciences biologiques et la médecine. Les expériences RMN peuvent être conçues efficacement pour sonder les angles de liaison, les distances entre noyaux, les constantes de couplage, les environnements électroniques, les constantes cinétiques, etc.

Relaxation des spins nucléaires

Le jeu d'expériences de relaxation qui est utilisé le plus couramment pour sonder la dynamique du squelette peptidique consiste en la mesure des vitesses de relaxation longitudinale R_1 et transversale R_2 de l'azote-15 , ainsi que la mesure de l'effet Overhauser nucléaire (NOE) à l'état stationnaire ^{15}N-{^{1}H}. Ces vitesses fournissent une information sur les fluctuations des interactions dipôle-dipôle ^{15}N-^{1}H et sur l'anisotropie du déplacement chimique ^{15}N. La fonction d'auto-corrélation de ces interactions n'est pas directement mesurée mais sa transformée de Fourier, la fonction de densité spectrale, est échantillonnée aux fréquences de transition du système de spin (par exemple la fréquence de Larmor des noyaux proton et azote-15).[23] La cartographie de la fonction de densité spectrale est la méthode la moins biaisée par un modèle pour décrire les mouvements du vecteur de la liaison H-N, fournissant ainsi des paramètres quantitatifs pour modéliser les mouvements globaux et locaux dans les protéines repliées et la distribution des mouvements dans les protéines désordonnées.[24,25]

Vitesses de relaxation paramagnétique

La vitesse de relaxation paramagnétique (paramagnetic relaxation enhancement, PRE) est proportionnelle à la moyenne $\langle r^{-6} \rangle$ de la distance entre un proton et un centre paramagnétique électronique. Le moment magnétique important de l'électron célibataire rend l'interaction dipolaire avec les spins nucléaires particulièrement forte. La détection de contacts longue distance, jusqu'à $\sim$ 25 Å de la sonde paramagnétique est ainsi possible.[26] La facilité d'implémentation des expériences de

PRE rend cette technique attractive pour sonder les contacts longue-distance dans les IDP en perturbant de façon minimale le système.

Figure 5: Le marquage des protéines sur un site spécifique avec une sonde paramagnétique permet de détecter directement des contacts longue distance à moins de ~ 25 Å de la sonde paramagnétique. Le PRE résultant des interactions magnétiques dipolaires entre un spin nucléaire et un électron célibataire d'un centre paramagnétique procure une information importante pour sonder les contacts longue distance entre différentes régions et segments d'une protéine qui sont éloignés dans la séquence mais proches spatialement.

Couplages Dipolaires Résiduels

En solution, les couplages dipolaires résiduels (RDC) peuvent être induits en alignant partiellement la molécule (mise en solution dans un milieu aqueux anisotrope) par rapport au champ magnétique, ce qui conduit à une moyenne incomplète de la distribution instantanée des orientations de la protéine. L'alignement partiel de la molécule fournit une information très utile sur la taille et la direction des vecteurs internucléaires.[27] En RMN la portée limitée du NOE ^{1}H-^{1}H jusqu'à environ 5 Å restreint l'analyse de l'information à un niveau local, et de ce fait ne permet pas de sonder des distances plus longues. A la différence des NOE, l'information contenue dans les couplages dipolaires est relative au champ magnétique, ce qui permet d'accéder à des orientations de vecteurs indépendantes de la distance. Les RDC fournissent ainsi des contraintes orientationnelles globales apportant une information fiable et importante complétant les NOE strictement locaux.[28] En raison de leur sensibilité à la conformation locale, les RDC ont été utilisés efficacement pour

décrire l'ensemble conformationnel des IPD.[29,30]

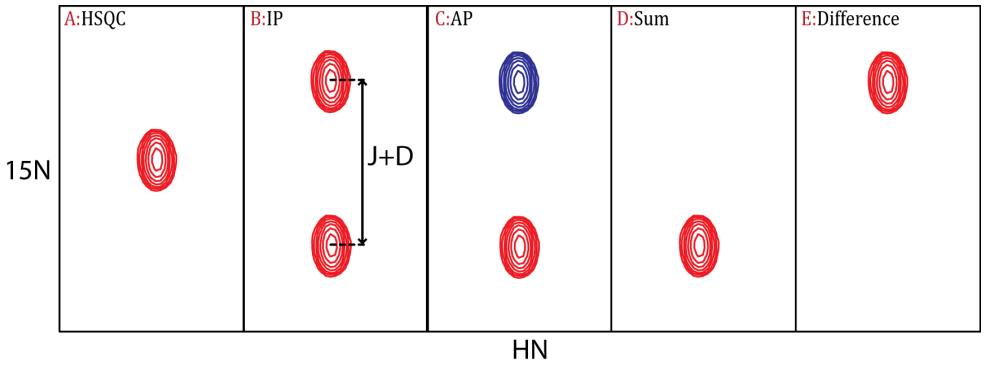


Figure 6: Principe de base des expériences IPAP. Deux expériences entrelacées sont enregistrées, correspondant aux composantes à haute et à basse fréquence. Les couplages sont extraits simplement en mesurant directement l'écart des pics (en Hz) dans les deux spectres séparés. A: pas de doublet (HSQC normal), B: doublet en phase dû aux couplages J+D (composantes ^{15}N à basse fréquence et à haute fréquence en solution anisotrope), C: doublet antiphase dû aux couplages J+D ; D : somme des spectres en phase et en anti-phase ; E : différence.

Etude des régions désordonnées d'Engrailed-2 basée sur des expériences RMN (expériences de relaxation, PRE et RDC)

La structure de l'homéodomaine d'Engrailed-2 de poulet a été déterminée par RMN. Elle montre le repliement typique de l'homéodomaine, avec un cœur à trois hélices et un bras N-terminal (Figure 17a).[31] La structure cristallographique d'un complexe entre l'homéodomaine d'Engrailed de drosophile et un ADN double brin (Figure 17b) indique que la troisième hélice α interagit dans le grand sillon de l'ADN tandis que le bras N-terminal interagit avec le petit sillon.[32] Le rôle des régions désordonnées dans la fonction de la protéine Engrailed-2 n'est que partiellement connu et, en particulier, son importance dans la modulation de la liaison à l'ADN demeure mal compris. La région désordonnée comprend plusieurs sites d'interaction connus avec d'autres partenaires ou cofacteurs protéiques. Notamment, la séquence hexapeptidique hydrophobe WPAWVY dans cette région constitue un site d'interaction pour d'autres homéoprotéines de la classe Pbx.

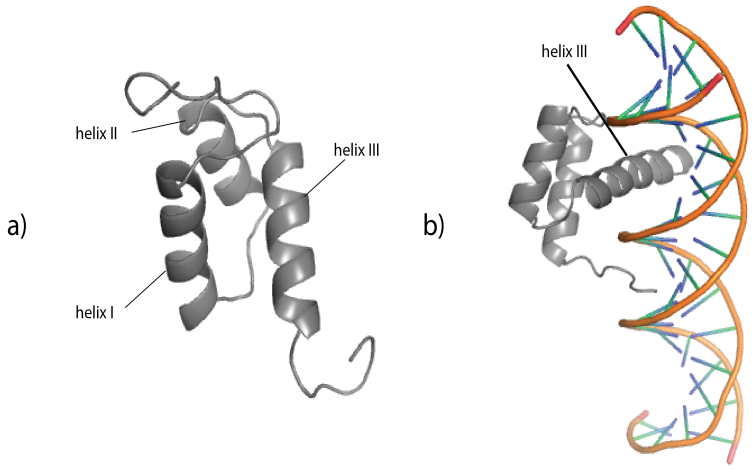


Figure 7: (a) Structure RMN de l'homéodomaine d'Engrailed 2 de poulet (code pdb 3ZOB) montrant les trois hélices alpha connectées par de courtes régions en boucles. (b) Structure cristallographique de l'homéodomaine d'Engrailed de drosophile en complexe avec un ADN double brin (code pdb 1HDD) montrant que la troisième hélice de l'homéodomaine est le site majeur de liaison à l'ADN.

La compréhension des fonctions importantes d'Engrailed-2 comme facteur de transcription nécessitent de caractériser les propriétés physico-chimiques moléculaires associées à une résolution atomique. Ce travail présente une analyse structurale, dynamique et fonctionnelle d'une construction d'Engrailed-2 de 114 résidus (146-259), appelée hdn, qui comprend l'homéodomaine étendu du côté N-terminal pour englober le site hexapeptide. La construction hdn d'Engrailed-2 comprend trois régions jouant un rôle fonctionnel important (Figure 18): l'homéodomaine impliqué dans la liaison à l'ADN et dans un mécanisme de translocation membranaire, le motif hexapeptide de liaison à d'autres protéines et un site de phosphorylation post-traductionnelle de sérines (contrôlant la sécrétion de la protéine).

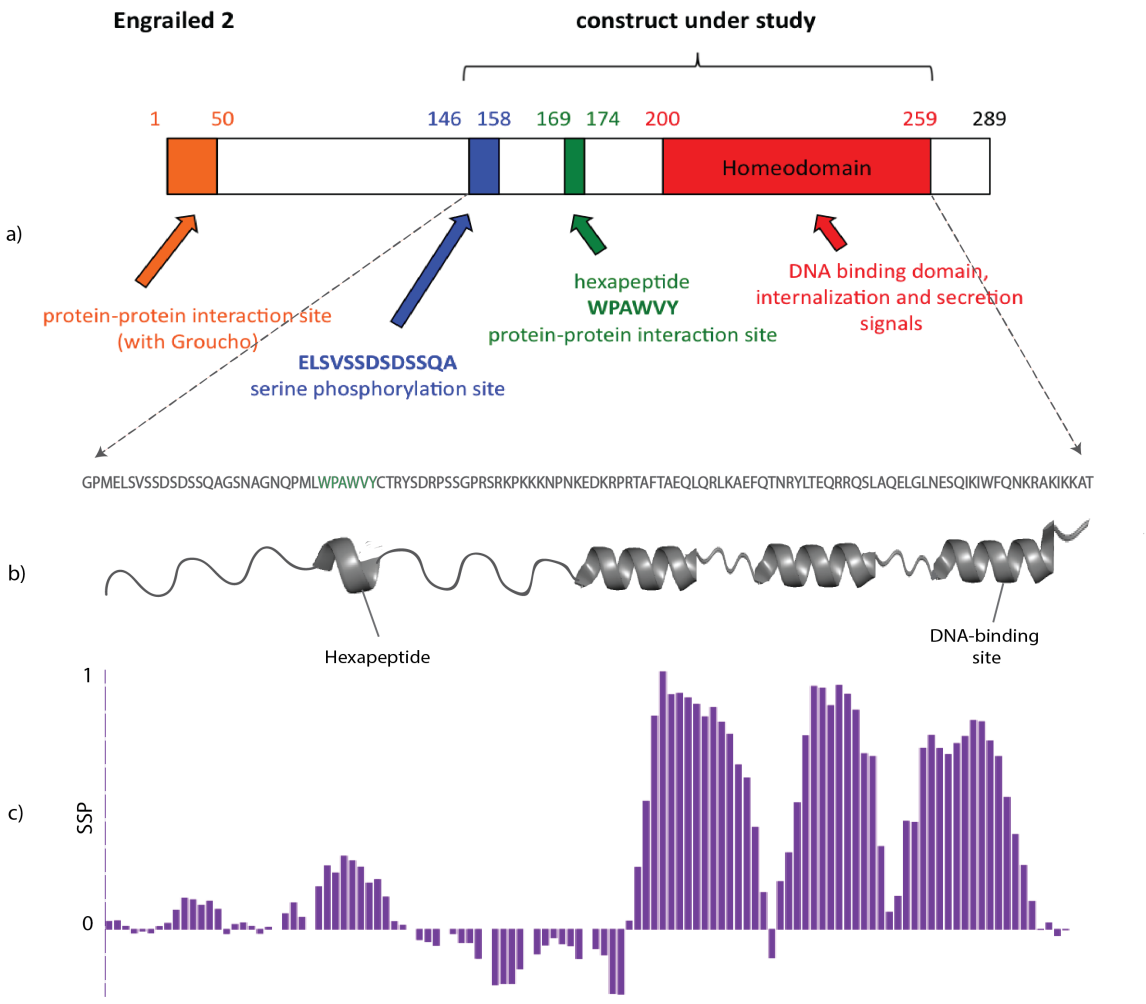


Figure 8: (a) Le facteur de transcription Engrailed-2 de poulet (289 résidus) possède un homéodomaine bien structuré (en rouge) et une longue région N-terminale de 200 résidus, intrinsèquement peu structurée, qui contient un site de phosphorylation (en bleu) et des sites de liaison à d'autres protéines (en vert et orange). (b) Représentation schématique de la structure secondaire de la construction étudiée hdn, montrant les trois hélices alpha de l'homéodomaine structuré, reliées par de courtes boucles, et l'hexapeptide dans la région désordonnée ayant une certaine propension à adopter une structure secondaire. (c) Score SSP de la construction hdn montrant clairement les trois hélices alpha de l'homéodomaine globulaire. L'hexapeptide a un score SSP supérieur aux autres régions désordonnées.

L'étude de la construction hdn devrait permettre de mieux comprendre les bases moléculaires et les fonctions biologiques associées. En particulier, un objectif de l'étude est d'analyser dans quelle mesure l'hexapeptide distant peut moduler la liaison à l'ADN de l'homéodomaine en l'absence d'interaction avec d'autres cofacteurs ou partenaires protéiques.

Distribution des mouvements pico- et nanoseconde d'Engrailed-2 à partir de la relaxation des spins nucléaires

Nous avons introduit une nouvelle approche nommée interprétation des mouvements par une projection sur une grille de temps de corrélation ("Interpretation of Motions by a Projection onto an Array of Correlation Times", IMPACT) pour analyser les données de relaxation à plusieurs champs dans les protéines désordonnées. La méthode est conçue pour dépendre le moins possible d'un modèle physique particulier des mouvements des protéines. Nous définissons un ensemble de temps de corrélation τ_i (et leurs fréquences réciproques $\omega_i = 1/\tau_i$) dans un intervalle qui est effectivement échantillonné par la relaxation ^{15}N. La fonction de densité spectrale expérimentale est alors reproduite comme une somme de fonctions lorentziennes $J_i(\omega)$, une pour chaque temps de corrélation τ_i selon l'équation :

$$J(\omega) = \sum_{i=1}^{n} J_i(\omega) = \frac{2}{5} \sum_{i=1}^{n} \frac{A_i \tau_i}{1 + (\omega \tau_i)^2}$$

où A_i est le coefficient du temps de corrélation τ_i dans la fonction de densité spectrale. Le résultat de ce processus, similaire à la projection sur une base de fonctions lorentziennes, est une distribution discrète de temps de corrélation dans un intervalle pertinent pour la relaxation. Le caractère multimodal de la distribution de temps de corrélation peut ainsi être mis en évidence, et les temps de corrélation les plus pertinents pour décrire les mouvements du squelette sont clairement révélés.

Nous avons appliqué IMPACT aux données collectées sur Engrailed 2 à cinq champs magnétiques différents et nous avons comparé les résultats avec deux modèles utilisés couramment pour lesquels la fonction de densité spectrale est représentée comme une somme de deux ou trois lorentziennes pour décrire les mouvements avec deux (2CT) ou trois (3CT) temps de corrélation, respectivement. Ces modèles sont identiques aux approches "model-free" et "model-free" étendu.

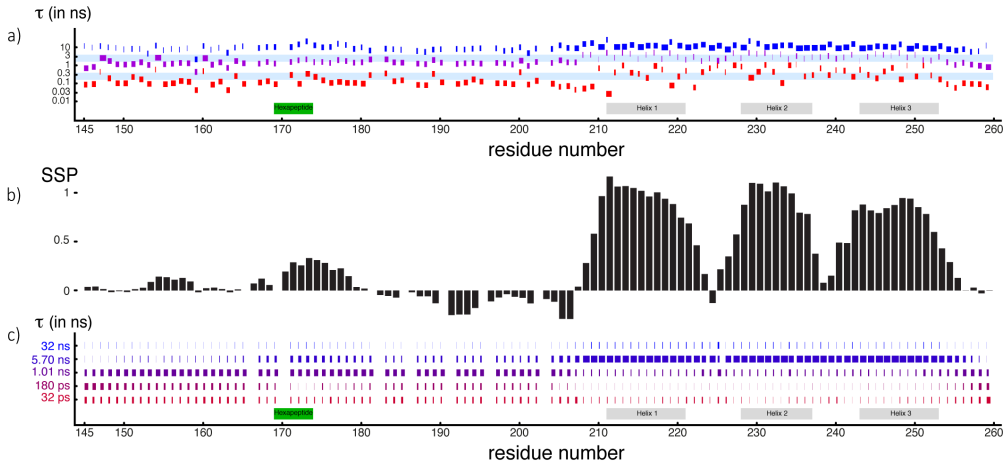


Figure 9: Représentations graphiques des analyses 3CT (a) et IMPACT (c) de la fonction de densité spectrale pour Engrailed-2. Les histogrammes sont tracés pour tous les résidus avec les règles suivantes : pour chacun des trois temps de corrélation $\tau_{a,b,c}$ déterminés en (a) et pour chacun des six temps de corrélation τ_i (i = 1, 2, …6) considérés en (c), un rectangle représente la dynamique de l'acide aminé correspondant le long de l'axe horizontal. La largeur de chaque rectangle est proportionnelle aux poids correspondants $B_{a,b,c}$ (a) ou A_i (c). En (a), les barres horizontales en bleu clair représentent la distribution des temps de corrélation τ correspondant aux fréquences réciproques entre $40 < 1/(2\pi\tau) < 100$ MHz ou entre $348 < 1/(2\pi\tau) < 870$ MHz. Les rectangles gris en (a) et (c) indiquent les hélices alpha, et le rectangle vert montre la localisation de la séquence hydrophobe hexapeptidique. (b) La propension à adopter des structures secondaires (SSP) est montrée pour faciliter la comparaison entre les caractéristiques structurales et dynamiques.

La représentation en code barre des coefficients IMPACT illustre parfaitement les variations de l'ensemble des temps de corrélation entre les éléments structuraux successifs : dans l'hexapeptide, par exemple, la diminution des mouvements dans la gamme inférieure à la nanoseconde est accompagnée par une augmentation dans la gamme au dessus de la nanoseconde. De façon similaire, la variation des coefficients pour les temps de corrélation aux extrémités N- et C-terminales de l'homéodomaine (résidus 200-210 et 254-260) illustre la transition continue des propriétés des mouvements le long de la séquence de la protéine. Enfin, même dans l'homéodomaine, où une analyse classique des données de relaxation devrait être la plus appropriée, les transitions dynamiques entre les hélices et les deux boucles sont clairement visibles et caractérisées quantitativement par l'approche IMPACT. La boucle α_1-α_2 présente une dynamique accrue à la fois dans les gammes de la nanoseconde et des dizaines de picosecondes, tandis que la boucle α_2-α_3 montre une augmentation significative mais plus modérée des mouvements. En conclusion, la représentation IMPACT fournit une vue élégante de la corrélation des propriétés

structurales et dynamiques, comme le montrent les scores SSP.

Les contacts longue-distance de la région désordonnée d'Engrailed-2 et leur contribution à la liaison à l'ADN

Les effets de relaxation paramagnétique ont été mesurés en comparant l'intensité des pics sur le spectre ^{1}H-^{15}N HSQC de l'état paramagnétique avec le spectre de référence ^{1}H-^{15}N HSQC dans l'état diamagnétique. Des contacts longue-distance multiples entre l'homéodomaine et la région N-terminale désordonnée ont pu être identifiés. Les effets de relaxation paramagnétique indiquent que la région désordonnée occupe un volume qui couvre des régions de l'homéodomaine dans au moins une partie de l'ensemble conformationnel d'Engrailed. Les résidus de l'homéodomaine ayant des PRE significatifs avec une sonde MTSL attachée dans la région N-terminale sont cartographiés sur la structure en solution de l'homéodomaine (code PDB 3ZOB)[31] (cf. Figure 46).

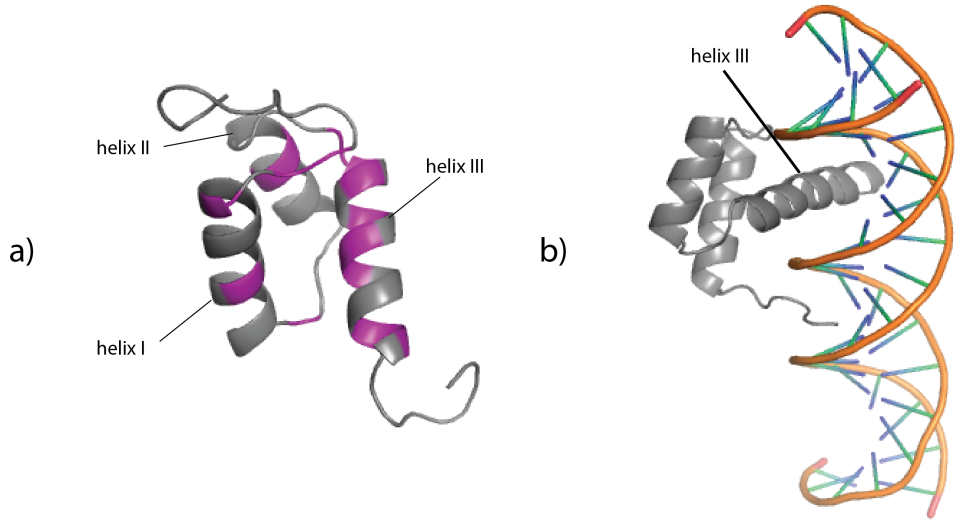


Figure 10: (a) Les résidus de l'homéodomaine les plus affectés (PRE < 50% induits par les contacts longue distance de la région désordonnée) sont représentés sur la structure RMN de l'homéodomaine d'Engrailed-2 (code pdb 3ZOB).[31] Les effets sur la troisième hélice indiquent la proximité de l'extension N-terminale dans des conformères ayant des populations significatives. (b) La structure cristallographique de l'homéodomaine d'Engrailed complexé à l'ADN résolue par Kissinger *et al.*[32] montre que la troisième hélice interagit avec le grand sillon de l'ADN (code pdb 1HDD).

Rôle de la région désordonnée dans la liaison à l'ADN

Les constantes de dissociation (K_d) de plusieurs constructions d'Engrailed-2 avec un ADN double brin ont été déterminées : construction hdn sauvage (incluant l'homéodomaine et la région désordonnée), homéodomaine avec un hexapeptide muté dans la région désordonnée (hdn_2W/A), homéodomaine sans hexapeptide (nls) et homéodomaine uniquement (hd). Les constantes de dissociation déterminées pour chacune des constructions d'Engrailed-2 sont résumées dans le tableau 2 ci-dessous.

Construct	**Direct Binding** K_d(nM)	**Competitive Binding** K_d(nM)
hdn_WT	177	18.91
hdn_2W/A	108	22.73
nls	104	10.8
hd	217	27.5

Tableau 1: Constantes de dissociation (K_d) des différentes constructions d'Enfrailed-2 testées pour un ADN marqué avec une sonde fluorescente.

Les constantes de dissociation indiquent clairement que toutes ces constructions lient l'ADN avec une bonne affinité. Ces résultats suggèrent une contribution possible des régions désordonnées pour moduler l'affinité de liaison à l'ADN. La plus grande affinité de la construction mutée nls (sans hexapeptide) par rapport à hdn suggère que l'hexapeptide exerce un effet d'auto-inhibition en l'absence d'autres cofacteurs protéiques.

Couplages dipolaires résiduels (RDCs) pour analyser l'espace conformationnel de la région désordonnée d'Engrailed-2.

Les couplages dipolaires résiduels à une liaison ^{1}H-^{15}N, C_α-H_α et C_α-C′ mesurés sur Engrailed-2 montrent une fluctuation des valeurs en fonction de la séquence. Les RDC mesurés dans l'homéodomaine d'Engrailed-2 montrent un degré assez important d'alignement. Les valeurs plus faibles de RDC mesurées dans la région désordonnée illustrent l'hétérogénéité conformationnelle conduisant à une moyenne de tenseurs d'alignement multiples dans la région désordonnée.[27] Les valeurs plus élevées de RDCs mesurées dans la partie désordonnée entre l'hexapeptide et l'homéodomaine suggère que cet IDR est plus ordonné que l'IDR situé du coté N-terminal de l'hexapeptide, en accord avec l''analyse IMPACT. La distribution des RDC ^{1}H-^{15}N le long de la séquence de l'homéodomaine montre un profil qui suit étroitement la structure secondaire. [33,34]

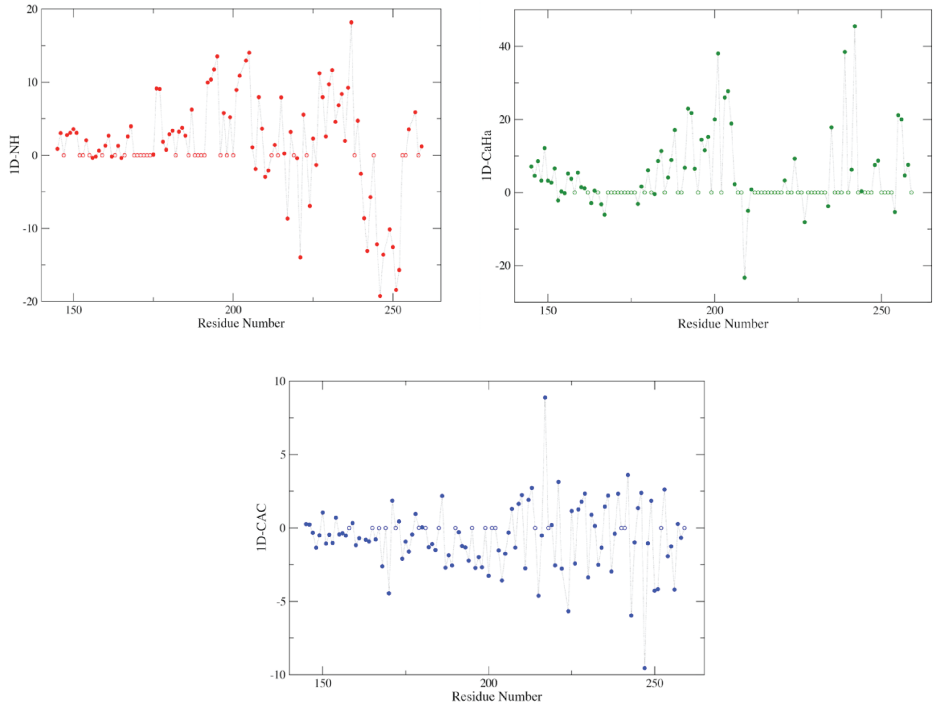


Figure 11: Valeurs mesurées pour les couplages dipolaires résiduels à une liaison ^{1}H-^{15}N, C_α-H_α et C_α-C′ de la protéine Engrailed-2 correspondant à la différence des écarts entre les doublets enregistrés dans l'état aligné (échantillon en gel) et l'état isotrope (échantillon en solution). Les ronds blancs correspondent à des données absentes en raison de pics peu intenses ou trop élargis dans l'état aligné.

Conclusion

L'approche IMPACT fournit une meilleure description quantitative de la fonction de densité spectrale dans les IDP/IDR, comme nous avons pu le mettre en évidence sur la protéine Engrailed. L'étude des PRE indique que des contacts transitoires longue distance existent entre la région désordonnée et l'homéodomaine structuré d'Engrailed-2, en particulier au niveau de la région impliquée dans la liaison à l'ADN. Les études d'interaction Engrailed-ADN mettent en évidence une contribution des régions désordonnées pour la liaison à l'ADN. Les RDCs mesurés indiquent un alignement important de l'homéodomaine et moindre de l'IDR de Engrailed-2 et illustrent l'hétérogénéité structurale et dynamique de l'IDR. Ces données serviront à définir plus précisément l'espace conformationnel d'Engrailed 2.

Chapter 1 Intrinsically Disordered Proteins

Chapter 1 Intrinsically Disordered Proteins

1.1 Introduction

Proteins are fundamental components of all living cells, required for structure, function, and regulation of the body's cells, tissues, and organs. Examples of proteins include enzymes, hormones, and antibodies that are essential for the proper functioning of an organism. The structural diversity of protein enables them to perform many diverse functions.

Amino Acid Sequence → Three Dimensional Structure → Protein Function

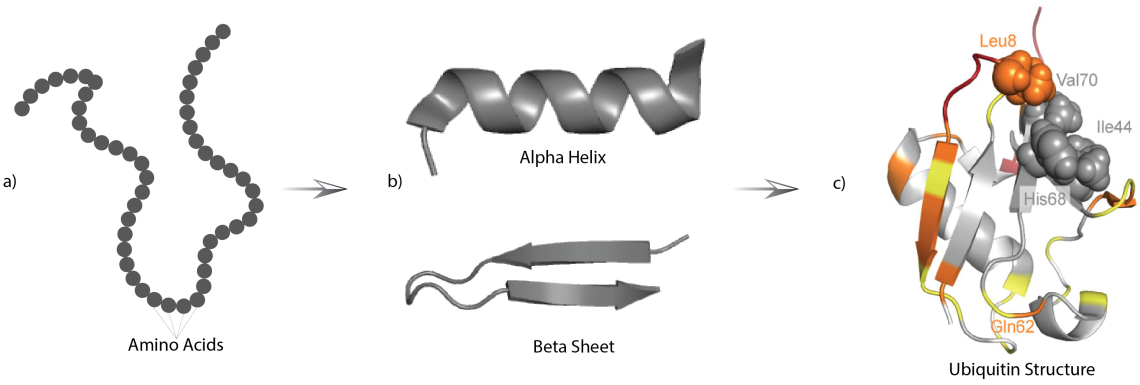


Figure 12: (a) Proteins are long chains of amino acids. (b) These one-dimensional polymer chains generally fold into an amazing variety of three-dimensional structures. (c) These 3D structures allow proteins to perform their wide range of functions.

The protein structure/function paradigm has been the central dogma of structural biology and other protein sciences over decades, stating that function strictly depends on prior 3D structure. The structure of a protein is stabilized by a variety of non-covalent intramolecular interactions such as hydrogen bonds, electrostatic interactions and Van der Waals forces. These interactions can be weakened and the protein structure can thus be destabilized by denaturants leading to protein unfolding.[1,2] Depending on the

denaturing conditions and the intrinsic properties of a protein diverse unfolded states can be observed, from partially unfolded states that retain a majority of the native structure to fully unfolded states that are close to a random coil owning neither alpha-helical nor beta-sheet secondary structure.[3,4]

More than 30 years ago, the first crack in the dogma requirement of the existence of a 3D structure for a protein to be functional appeared with the discovery of protein segments now called Intrinsically Disordered Proteins (IDPs) or Intrinsically Disordered Regions (IDRs) lacking a stable 3D structure but yet fully functionnal.[5] Computational analysis shows that estimates of wholly disordered proteins in the bacteria ranged from 1% to 8%. The eukaryotic cells have 33% of proteins that contain disordered segments > or = 30 amino acids in length and 31% of humans proteins are 35% unstructured.[35'16]

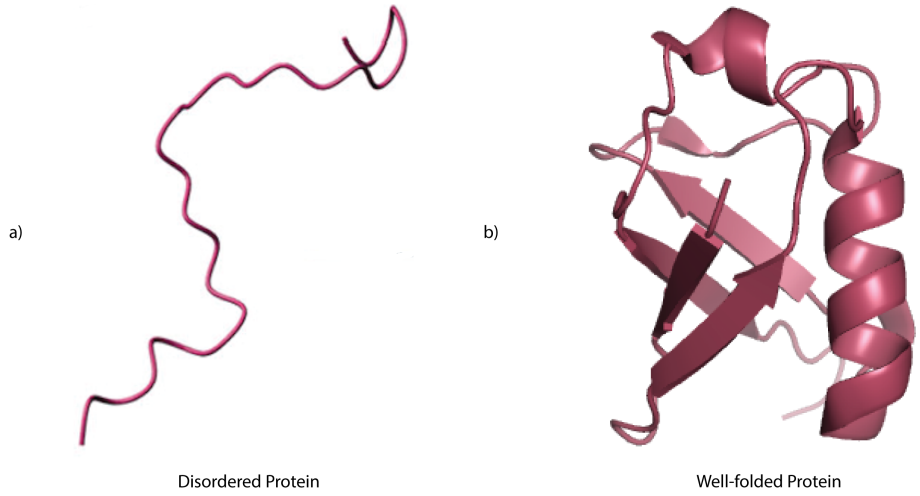


Figure 13: (a) Protein lacking a fixed or ordered 3D structure called IDPs. (b) Protein with a well-defined 3D structure (Ubiquitin; pdb code 1d3z).

The ubiquitous existence of IDPs/IDRs, which lack a well-defined three-dimensional structure but nevertheless play crucial roles in various cellular processes, has not only challenged the structure–function paradigm but also has opened a new class of intriguing substrates for studies in modern proteomics.[6,7] IDPs and IDRs are involved in a wide variety of biological processes, from regulation, cell signaling and control and thus well complement the functions of ordered proteins. The IDPs p21 and p27 bind to cyclin-dependent kinases (cdks) and thus function as a regulator in mammalian cell

cycle through direct interactions with Cdk/cyclin complexes.[8] The transcription factor Myc that binds to DNA to regulate the cell division in many cell types is intrinsically disordered alone and becomes folded upon binding to DNA.[9] Alpha-synuclein, a IDP in unbound state is potentially involved in synaptic rearrangement and also acts as a molecular chaperone in synaptic activity.[10] Many human diseases, including cancer, amyloidosis, diabetes, cardiovascular and neurodegenerative diseases are associated with the misidentification and missignaling of some IDPs.[11] The development of novel strategies for drug discovery targeting IDPs is the consequence of their involvement in these diseases.[12]

1.2 Classifications of intrinsically disordered regions and proteins

The classification schemes of structured proteins, which define families of homologous sequences cannot explicitly accommodate the nature of IDRs/IDPs, since the definition of homologous regions in the absence of structural constraints is an open challenge.[36,37] The need to classify disordered regions in proteins aims to predict and understand their function. Their classification requires an accurate integration of information related to their sequences, structures, interactions, functions and biophysical properties. The classification of IDRs/IDPs based on their diverse functions due to their highly flexible nature exemplifies the functionalities of IDRs/IDPs. We can classify IDPs and IDRs into different categories owning to different functional classes as described below.[16]

1.2.1 Entropic Chains

Entropic chains are IDRs/IDPs that carry out their functions with almost no restriction of the conformational space (besides steric repulsion) and maintain such a high degree of disorder throughout the biological processes they are involved in, e.g. flexible linkers and spacers. The flexible linkers provide spatial freedom to protein domains that are positioned at either side of the linkers, while maintaining the cooperative inter-domain interactions.[16] Spacers can regulate the distances between protein domains and provide spacing so that two adjacent domains do not sterically interfere with one another.[38]

1.2.2 Display Sites

The highly flexible nature of IDRs/IDPs helps in stabilizing the crucial steps involved in protein biosynthesis within the cell. In an IDP or IDR, the chain is generally exposed and its structural plasticity makes it a favored substrate for enzymes, in particular those in charge of executing post-translational modifications (PTMs). IDPs and IDRs play essential roles in PTMs, which regulate protein function within the cell. The flexibility of IDRs/IDPs facilitates not only the binding to enzyme active sites but also the exposure of PTM motifs for efficient signalling.[39]

1.2.3 Chaperones

Chaperones are functionally related group of proteins that stabilize unfolded proteins and assist RNAs and proteins to fold properly in the cell under physiological conditions as well as under stress.[40] About 50% of the protein chaperones of RNA are composed of disordered regions while protein chaperones are made up of about 33% of disordered regions.[41] The flexibility and variety of side chains (i.e. polar, acidic, basic, hydrophobic) of disordered segments of chaperones enable them to interact with a wide variety of polypeptide chains and increase the association rate of the interaction.[42] The chaperones mediated folding of proteins into their native tertiary structures reduces the amount of unfolded proteins with a loss of function and prevents the formation of toxic aggregates.

1.2.4 Effectors

Effectors are IDRs/IDPs that show disorder-to-order transition upon binding to other proteins. The folding upon binding mechanics allows effectors to modify the activity of their partner proteins especially in cellular signal transduction cascades e.g., p21 and p27 that regulate different cyclin- dependent kinases (Cdk).[18] Effectors can also positively or negatively regulate the same protein through their competitive interaction to the active or binding sites of the protein. For instance, the disordered part of the LAB homeoprotein reduces the affinity of the homeodomain for DNA in the absence of a partner protein. The presence of synergetic partner (EXD) protein allows the

homeodomain of LAB to bind DNA with high affinity and specificity in a ternary complex.[43]

1.2.5 Assemblers

These are the IDRs/IDPs that display several binding sites to multiple partner proteins and promote the formation of higher-order protein complexes e.g., ribosome.[44] The wide conformational space of IDRs/IDPs allows them to efficiently reach the binding surface of different protein partners and gather them within a molecular machine with minimal steric clash. Assembled high order complexes can then fulfill essential roles in the cell.[16]

1.2.6 Scavengers

These are the IDRs/IDPs that participate to store and neutralize small ligands or foreign substances with high ligand specificity. For instance, the IDP Chromogranin A stores adrenaline and adenosine triphosphate in adrenal gland.[45] NMR analysis of Chromogranin A shows that it is highly disordered in both the free and bound form.[45] Casein is another example of a scavenger that initiates the solubilization of clusters of calcium phosphate in milk and other biofluids.[46]

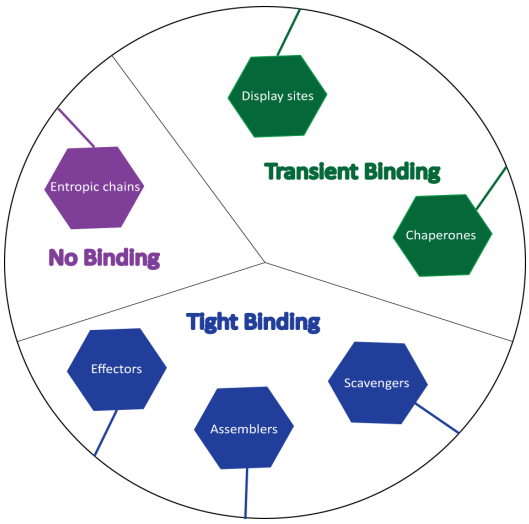


Figure 14: Classification of IDRs based on their functions and the nature of their interactions.[16]

1.3 Structural Heterogeneity of IDPs

The structural plasticity and vast conformational space of IDPs/IDPRs empower them to interact with many binding partners and thus engage efficiently in transient regulatory networks.[17] The ability of these proteins to gain secondary or tertiary structural elements upon binding to biological partners keeps them puzzling in many aspects of their structure and dynamics.[18] Their heterogeneous assembly of conformations ranging from extended conformers to compact structures forces them to be conceived as "ensembles of a continuum of rapidly interconverting structures". Interestingly, a combined study of the IDP osteopontin by Electron Paramagnetic Resonance (EPR) and Nuclear Magnetic Resonance (NMR) shows that some structural elements are stabilized by long-range interaction patterns similar to those found in folded proteins. This combined investigation of osteopontin demonstrates that its both extended and cooperatively folded states are populated in its apo form.[13] The characterization of the conformational space and quantitative description of the dynamics of IDPs at atomic level aims to understand their biological function.

X-ray crystallography, the major protein structure determining technique, is not suited for structural characterization of IDPs. The missing electron densities for commonly unobserved atoms in proteins that fail to scatter X-rays coherently are generally believed to be disordered. The small-angle-X-ray scattering (SAXS) can distinguish between folded and disordered proteins, which determines the shape and size of the proteins by measuring the intensity and the scattering pattern of the scattered X-rays by the sample in solution state.[19] The single-molecule fluorescence resonance energy transfer (smFRET) can provide information about protein structure and dynamic conformational changes at single molecule level.[20] Atomic force microscopy (AFM) is also useful for the observation of structural dynamics of proteins and can probe the conformational heterogeneity in real time.[21]

The extent of disorder is not protein specific; even a single amino acid in overall static protein can fluctuate to various degrees of structural heterogeneity. This structural and dynamical heterogeneity of IDPs makes NMR a prime method to investigate the conformational space and the kinetics of its exploration at atomic resolution.[22] NMR

spectroscopy is a powerful structural biology technique. NMR is highly adapted to the study of IDPs and IDRs as it unique in its ability to provide information at the resolution of a single atom. The wealth of information contained in chemical shifts, relaxation rates, scalar and dipolar couplings and paramagnetic relaxation enhancements (PRE) makes of NMR the tool of choice to investigate the conformational and dynamical properties of IDPs.

The present work aims at characterizing the conformational space and the timescales of its exploration in the disordered region of Engrailed-2 transcription factor. We have used ^{15}N relaxation rates and developed a new approach to quantify the distribution of correlation times of motions in IDPs. We have also aimed at describing the conformational space of Engrailed 2. In particular, long-range contacts of the disordered region with the DNA-binding globular domain have been explored and their contribution in the binding of Engrailed-2 to DNA have been quantified.

Chapter 2 Intrinsically disordered protein Engrailed-2

Chapter 2 Intrinsically disordered protein Engrailed-2

2.1 Homeoproteins

Homeoproteins constitute a large family of transcription factors that regulate the expression of specific target genes and are active throughout development and in adulthood.[47,48] These proteins consist of a highly conserved 60-residue domain called homeodomain which comprises a short extended strand followed by three alpha helices connected by short loop regions.[32] The homeodomain bears most of the DNA-binding affinity of the homeoproteins to regulate the transcription of genes.[49] In addition to their role as nuclear transcription factors, homeoproteins have the amazing property to be secreted and internalized using unconventional pathways involving direct translocation through cell membranes. This transduction ability to pass from cell to cell without subjecting to classic endocytosis mechanism has been showed to be involved in the regulation of neuronal morphology and axon guidance. Once internalized homeoproteins do not only act as transcription factors but also regulate translation. Their role as translation regulators is mediated by a conserved sequence that requires a direct interaction with eIF4E factor to initiate the translation process. The membrane translocation is a specific property inherited by the conserved homeodomain of these homeoproteins. The mechanism of their secretion involves the phosphorylation of a serine-rich region located outside of the homeodomain.[50]

Different homeoproteins can interact together to form ternary complexes with DNA. The molecular basis for the recognition between homeoproteins was studied by X-ray crystallography. The crystal structure of a ternary complex of the homeoproteins Ultrabithorax (Ubx) and Extradenticle (Exd) with DNA solved at 2.4 Å resolution shows that the homeodomains of Ubx and Exd bind opposite faces of the DNA (Figure 15a).[51] A similar structural arrangement was found for homeoproteins HoxB1–Pbx1 bound to DNA (Figure 15b).[52] In both cases a conserved motif called hexapeptide in the disordered region of Ubx or HoxB1 binds in the pocket of partner proteins Exd or

Pbx1 respectively. The heterodimer formation mediated by the hexapeptide leads to an enhancement of the binding specificity to DNA target sequences. These observations demonstrate the general principle of interaction of homeoproteins, as evidenced for Ubx/Exd and HoxB1/Pbx1, and clearly emphasize the role of the hexapeptide motif in cooperative interactions between homeoproteins.

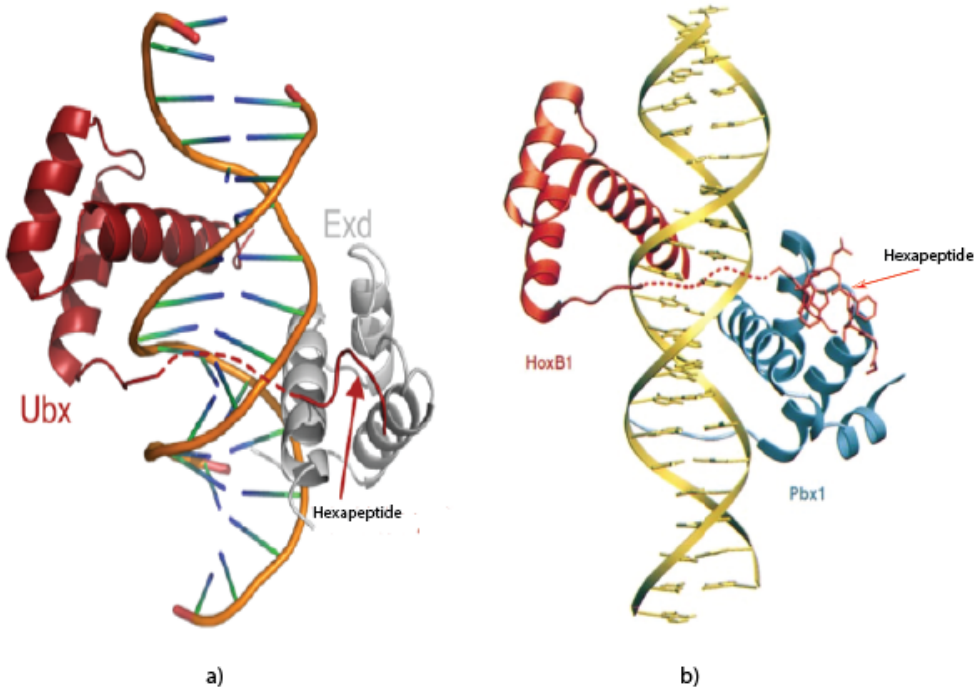


Figure 15: An overview of the (a) Ubx-Exd-DNA and (b) HoxB1-Pbx1-DNA ternary complex. In both cases homeodomains bind on opposite faces of the DNA. The dashed red line represents the disordered linker between the homeodomain and the Exd/Pbx binding site called hexapeptide (consensus sequence Y/F-P-W-M-K/R). The hexapeptide binds in a pocket on the surface of the Exd and Pbx1 homeodomains, respectively.

In addition to this well-documented interaction, the hexapeptide could also be implicated in a *direct* modulation of homeodomain DNA-binding properties. Indeed, a few data from the literature suggest that the hexapeptide and nearby disordered regions may inhibit homeodomain DNA binding in the absence of other interaction partners or cofactors. Chan and coworkers reported that mutations in the hexapeptide region of LAB homeoprotein enhances its DNA binding affinity in the absence of any other protein partners.[43] Another study from Wütrich's and Gehring's

groups on Antennapedia homeoprotein showed that the 60 residue homeodomain has a stronger affinity for DNA than a elongated 81 residue construct encompassing the N-terminal hexapeptide motif.[53]

2.2 Engrailed-2

Engrailed homeoproteins are encoded by two different genes in humans and higher organisms, engrailed 1 and 2, which exert slightly different functions. The physiological role of Engrailed in embryonic development and in the survival of dopaminergic neurons in adults prevails the importance of Engrailed in many multicellular aspects.[54] The role of Engrailed to initialize or repress the transcription processes depends on the presence of different transcription repressors or activators. Engrailed alone acts as repressor for polyhomeotic gene in *Drosophila* but the presence of Extradenticle homeoprotein (Exd) potentiates Engrailed to act as transcription activator.[55] The interaction of other protein cofactor with Engrailed also modulates the binding affinity of its homeodomain to the DNA. These interactions lead either to an increase or a decrease in the binding affinity of the homeoprotein to the DNA respectively.[32,53]

The sequence alignment of human, chicken and mouse Engrailed 1 and 2 proteins shows (Figure 16) that the homeodomain sequence is highly conserved in vertebrates. Highly conserved segments also exist in the disordered tails that span up to 50 residues (50-60, 165-185, 260-280) in the N- and C-terminal regions of the protein and act as interaction sites for partner proteins. We focused our work on Engrailed-2 from chicken. This protein has a slightly shorter length than its mammal counterparts and is well expressed as a recombinant protein in E. coli. Chicken Engrailed-2 comprises 289 residues inheriting a fully folded homeodomain (200-259) and disordered N- and C-terminal tails.

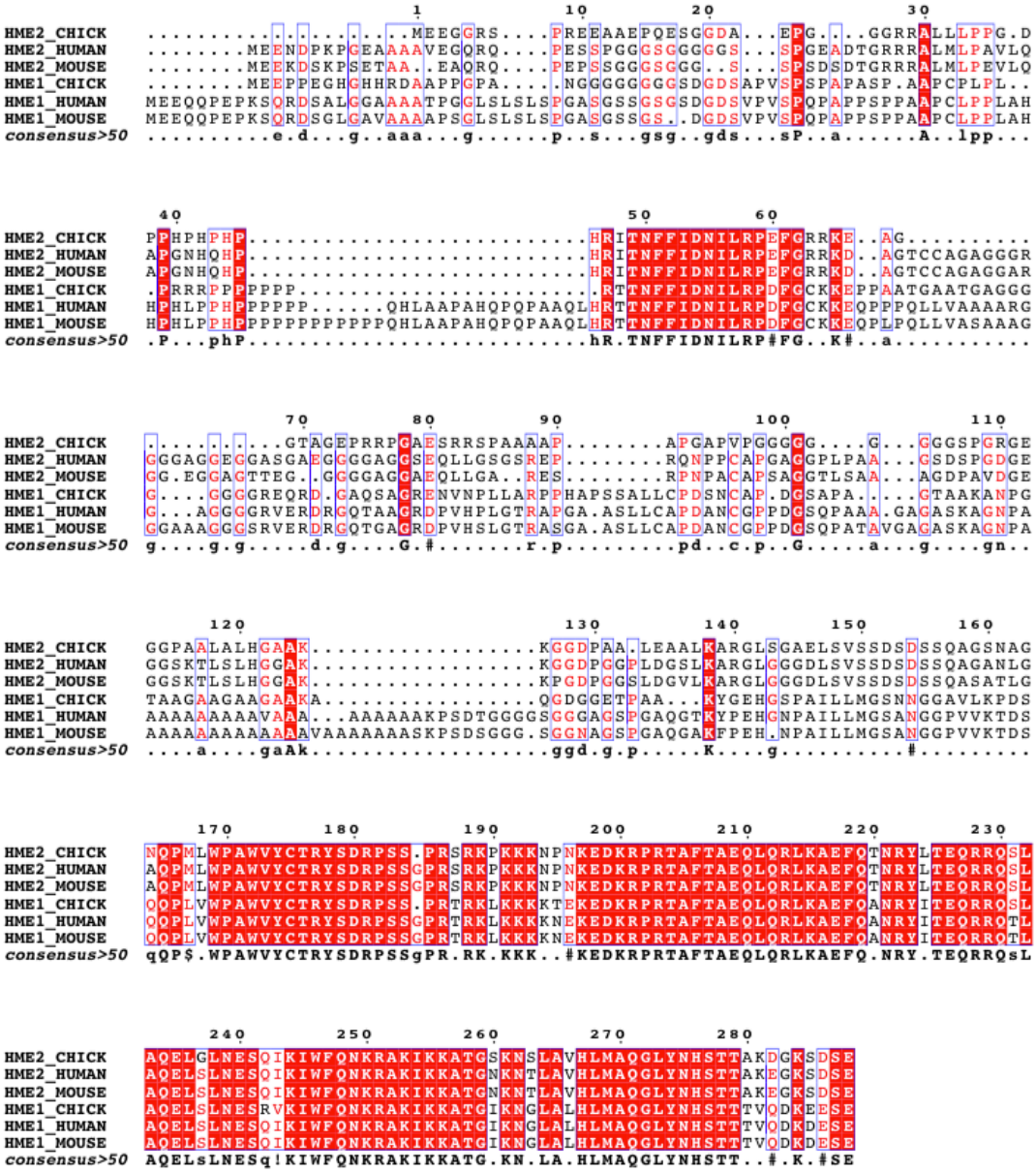


Figure 16: Sequence alignment of human, chicken and mouse Engrailed 1 and 2 (HME1 and HME2 respectively) showing the sequence similarity in homeodomain (region ~200-260) and other regions.

The structure of chicken Engrailed-2 homeodomain determined by NMR shows the typical homeodomain fold with a tri-helical core and an amino-terminal arm (Figure 17a).[31] The crystal structure of a complex containing the *Drosophila* Engrailed homeodomain and a duplex DNA site (Figure 17b) shows that the third α helix fits into the major groove while the N-terminal arm fits into the minor groove.[32]

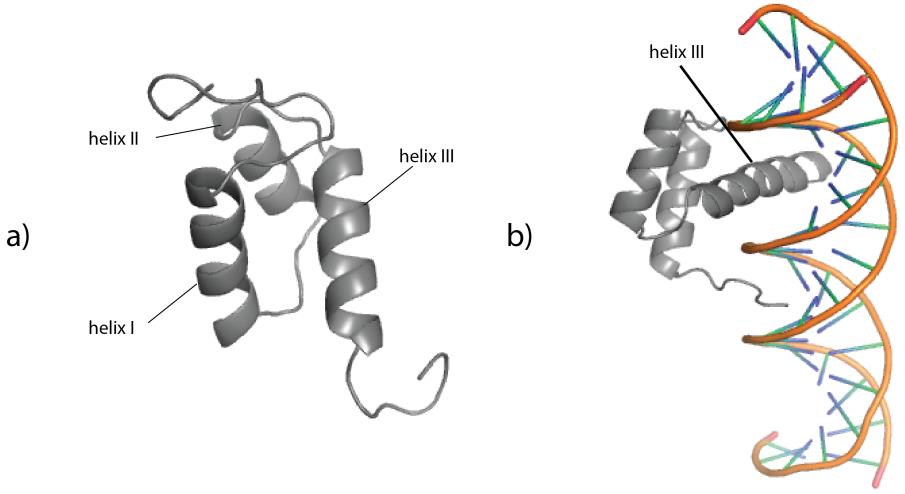


Figure 17: (a) NMR structure of chicken Engrailed 2 homeodomain (pdb code 3ZOB) comprising three alpha helices connected by short loop regions. (b) X-ray structure of Drosophila Engrailed homeodomain in complex with duplex DNA (pdb code 1HDD) showing the third helix of the homeodomain as a major binding site for DNA.

The role of the disordered region in the function of Engrailed-2 protein is only partially known and, in particular, its importance for the modulation of DNA binding remains unclear. The disordered region comprises known interaction sites for other partner proteins or cofactors, in particular the highly hydrophobic hexapeptide sequence (WPAWVY) in this region that acts as an interaction site for other homeoproteins belonging to the Pbx class.

The important functions of Engrailed-2 as a transcription factor and structural insight to perform these functions invoke to understand further associated processes at an atomic resolution. The present work adds structural, dynamical and functional insight of a 114 residue (146-259) Engrailed-2 construct (hdn) that comprises the homeodomain elongated on the N-terminal part to include the hexapeptide region. The study of hdn will help to understand the molecular mechanisms and their underlying biological functions. In particular, one aim of this study is to analyze how the distant hexapeptide region may modulate homeodomain DNA binding in the absence of interaction partner or cofactors.

This Engrailed-2 hdn construct contains three important functional regions (Figure 18): the homeodomain involved in DNA-binding and membrane translocation, a

protein-binding site (hexapeptide) and a serine phosphorylation site (controlling protein secretion). Prior NMR investigations in the laboratory have been carried out to assign ^{1}H, ^{15}N and ^{13}C resonances and to analyze the structure of this construct. In particular SSP (Secondary structure propensity) score calculated from C^{α}, C^{β}, CO and H^{α} chemical shifts clearly reveals that it comprises both fully folded and disordered regions.[56]

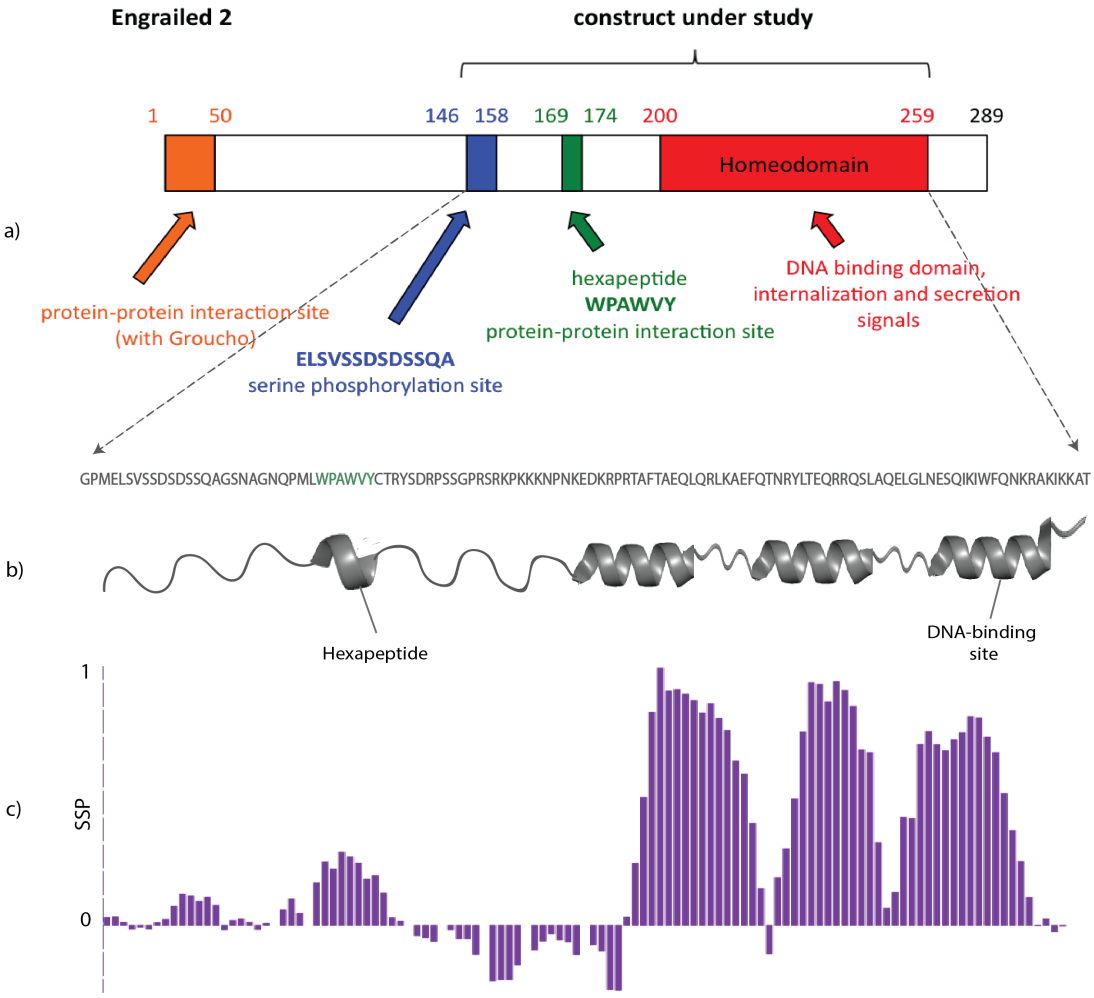


Figure 18: (a) Chicken Engrailed-2 transcription factor (289 residues) possesses a well-folded homeodomain (in red) and a long, 200-residue, mostly unstructured N-terminal region that contains a phosphorylation site (in blue) and protein interaction sites (in green and orange). (b) Schematic picture of the construct (hdn) under study showing the well-folded homeodomain comprising three alpha helices connected by short loop regions and hexapeptide in disordered region that has some secondary structure propensity. (c) SSP score of hdn construct clearly indicating the well-folded homeodomain comprising three alpha helices while hexapeptide has a SSP score higher than other disordered region.

Chapter 3 Expression and Purification of Engrailed-2

Chapter 3 Expression and Purification of Engrailed-2

3.1 Expression and Purification Protocol

Glutathione *S*-transferase (GST) tagged Engrailed-2 was expressed as a fusion protein in *E. coli* cells in order to facilitate the purification process. The retention of GST fused protein by using chromatographic columns helps to purify only the protein of interest. The protein fragments then can be separated by use of another round of chromatographic column after the cleavage of GST tag from protein. Reverse phase HPLC further helps to remove any impurity that is retained during the chromatographic purification process.

3.1.1 Transformation of the *E* coli BL21RP strains

The following protocol was adopted for the transformation of pGEX plasmid (Pharmacia) containing the gene coding for sequence 146-259 of chicken Engrailed-2 fused to GST. After taking out from -80 °C, the competent BL21RP cells were kept at 4 °C for 15 minutes. 2 µl (100 ng/µl) of plasmid was mixed with 20 µl of competent cell glycerol stocks and incubated at 4 °C for 10 minutes. The mixture was then subjected to 42 °C for 90 seconds for heat shock. After heat shock, 10 µl of LB was added and cells were incubated for 10 minutes at 4 °C. 5 µl of these transformed cells were spread on agar plate containing ampicillin and chloramphenicol at a concentration of 50 µg/ml and incubated overnight at 37 °C. Next morning, a single colony was picked up from agar plate and transferred to 50 ml of the LB medium containing ampicillin and chloramphenicol (100 ng/µl). The cells were then grown to $OD_{600} \approx 0.6$ and glycerol stock was prepared by mixing 300 µl of the culture with 200 µl of 50% glycerol and stored at -80 °C.

3.1.2 Expression of the recombinant Engrailed-2

Uniformly labeled (^{15}N or/and ^{13}C) Engrailed-2 protein was expressed and labeled using the Marley protocol.[57] A preculture was grown overnight at 37 °C by mixing 100 µl of glycerol stock (transformed cells) in 100 ml of LB medium supplemented with 100 µg/ml of ampicillin and chloramphenicol. Next morning, 10 ml of this preculture was added to 500 ml of LB medium in a 2 L flask (8 flasks, total culture volume of 4 L). The cells were grown at 37 °C to $OD_{600} \approx 0.65$ and centrifuged for 10 minutes at 4000 g for 15 minutes. The cell pellet was resuspended in 1 liter of M9 medium. After 15 minutes, $^{15}NH_4Cl$ or/and ^{13}C-glucose were added as nitrogen and carbon source for uniform ^{15}N or/and ^{13}C labeled protein. After 30 minutes, protein expression was induced by addition of 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). After 4 hours of induction, the cells were centrifuged at 4000 g for 15 minutes. The cell pellet was resuspended in 50 ml of phosphate buffer (pH = 7.4) containing 140 mM NaCl, 1 mM EDTA, 1 µg/ml pepstatin, 1 µg/ml leupeptin, 1.5 mM PMSF, 2 mM TCEP, 0.5% v/v Triton X-100, 100 µg/ml lysozyme and stored at -80 °C.

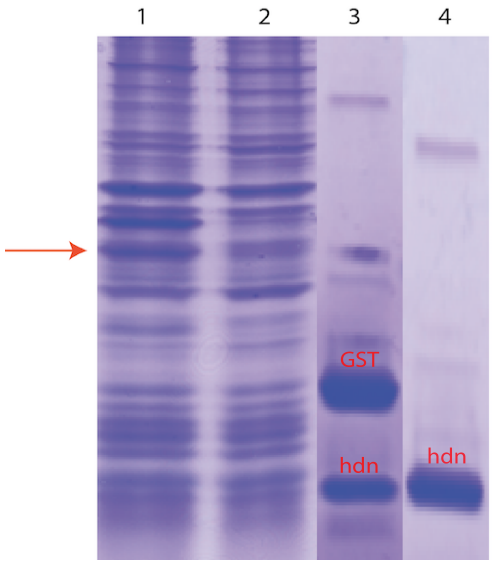


Figure 19: SDS page showing the different expression and purification steps involved in Engrailed-2 NMR sample preparation. Lanes 1 and 2 correspond to bacterial extract after and before induction with IPTG, respectively. Lane 3 shows separate bands for GST and hdn after cleavage with PreScission protease. Lane 4 shows the complete removal of GST after heparin column yielding purified hdn.

3.1.3 Purification Protocol of Engrailed-2

The cell pellets were thawed at room temperature and subjected to sonication for cell lysis. Cell debris were spun down at 20,000 g at 4 °C for 45 minutes and supernatant was filtered with 0.2 µm pore size filter. A GSTrap FF column (GE Healthcare) was equilibrated with PBS and supernatant was loaded at a flow rate of 0.5 ml/min to entrap GST-Engrailed fusion protein. The column was then washed with PBS followed by Tris-HCl buffer (50 mM, 1 M NaCl, pH = 7.5) to remove any unspecifically bound protein and DNA from cell debris. The GST-Engrailed retained to column was then eluted using 50 mM Tris-HCl buffer at pH = 8, with 150 mM NaCl, 1 mM EDTA, 1 mM DTT and 30 mM glutathione (pH = 8).

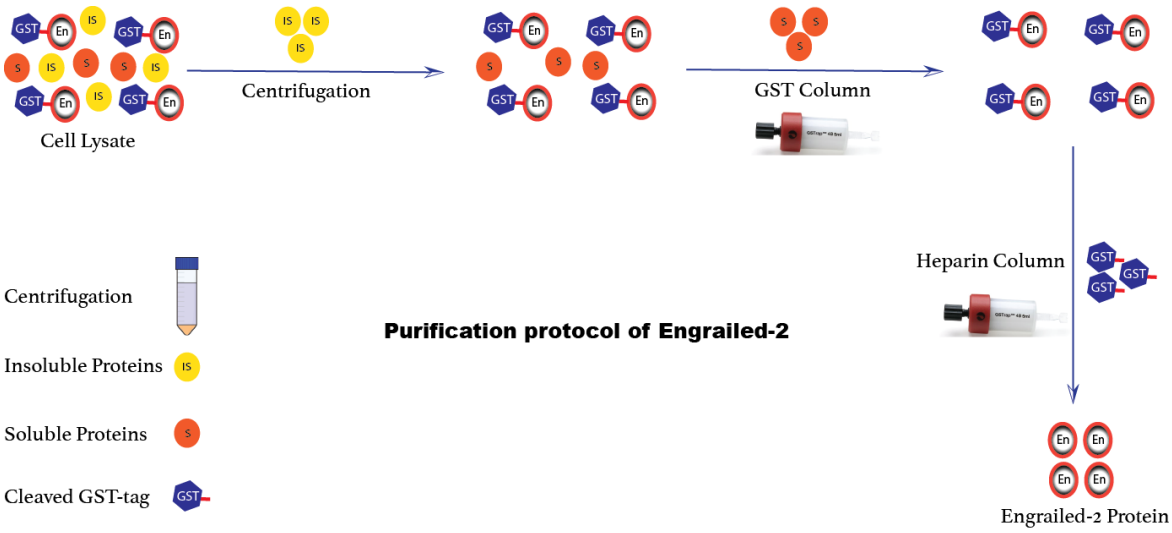


Figure 20: The different steps involved in purification of Engrailed-2. Centrifugation of cell lysate allows insoluble proteins to form a pellet that can be easily separated from supernatant containing Engrailed-2 and other soluble proteins. GSTrap FF column retains only GST tagged Engrailed-2 protein while passing all other soluble proteins to flow through. PreScission protease cleaved Engrailed-2 and GST tag can then be separated by HiTrap Heparin column that retains only the positively charged Engrailed-2 while leaving GST tag to pass through. After these steps Engrailed-2 can be further purified by gel filtration or reverse HPLC.

Extensive dialysis was carried out to remove glutathione before adding PreScission protease (2 units/100 µg of fusion protein) for GST tag cleavage. The cleavage took place overnight, resulting in three additional residues (G-P-M) at the N-terminal site of Engrailed. The cleavage was checked by SDS PAGE. The cleaved Engrailed-2 protein was then loaded on HiTrap Heparin column (GE Healthcare) to separate it from GST

tag. The net positively charged Engrailed-2 binds to negatively charged Heparin while GST tag has no affinity for Heparin column. Engrailed-2 was then eluted by using 50 mM sodium phosphate buffer, pH 7.4, containing 1 M NaCl at a flow rate of 1 ml/min.

The eluted fractions of Engrailed-2 were then purified by reverse phase HPLC using a Nucleosil® (Macherey-Nagel) C4 column (250 x 10 mm, 300 Å pore size, 5 µm particle size). The elution of the protein was achieved by using a 20 - 80% acetonitrile gradient in water, 0.1% TFA, at a flow rate of 4 ml/min. The Engrailed-2 protein was lyophilized and then dissolved in 40 mM succinate buffer pH = 6.0 supplemented with antiproteases (1 µg/ml leupeptin, pepstatin and AEBSF, 1 mM EDTA). 10 mM TCEP was added to prevent oxidation of cysteine.

3.2 Mass Spectrometry Analysis

The Matrix-assisted laser desorption/ionization coupled to time-of-flight mass spectrometer (MALDI-TOF) technique was used to analyze the molecular weight of Engrailed-2 protein. The observed m/z for ^{15}N labeled Engrailed-2 (hdn) was 13622 ± 20 Da, which was close to the theoretical estimate (13637 Da) for uniformly 100% labeled ^{15}N labeled protein.

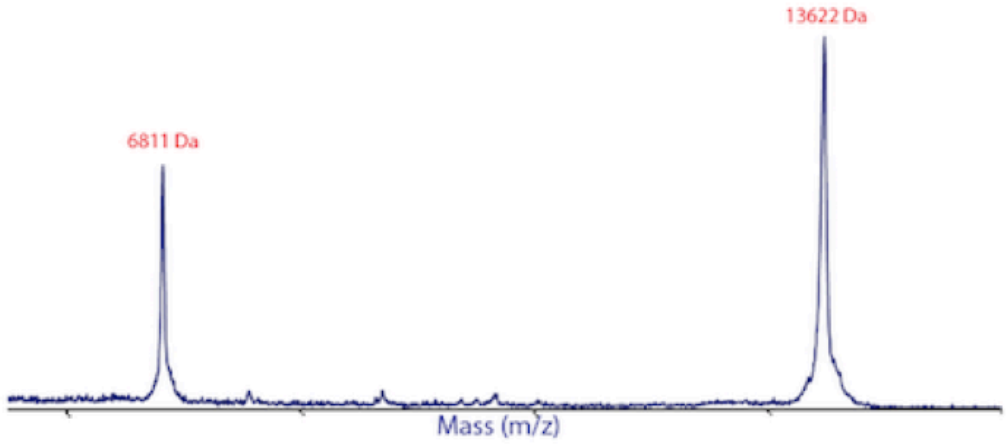


Figure 21: Observed m/z of ^{15}N labeled Engrailed-2 = 13622 Da (double charged peak = 6811 Da). The mass spectrum clearly indicates that Engrailed-2 purification protocol is highly efficient to yield Engrailed-2 protein without any contamination and degradation of the protein.

The same protocol was used for the expression and purification of all Engrailed-2 constructs used in this study as shown in Figure 22.

Figure 22: Different Engrailed-2 constructs used to study various aspects of Engrailed-2 function. All constructs contain three additional residues (pink) at the N-terminal part belonging to the PreScission cleavage site of Engrailed-GST fusion complex. The position of the cysteins is highlighted in green while hexapeptide is marked in blue.

Chapter 4 Distribution of pico- and nanosecond motions in Engrailed-2 from nuclear spin relaxation

Chapter 4 Distribution of pico- and nanosecond motions in Engrailed-2 from nuclear spin relaxation

4.1 Nuclear Magnetic Resonance (NMR)

NMR is an extremely powerful technique for investigating the structure and dynamics of biomolecules, in particular in solution, which provides physiological-like conditions. Samples of biomolecules enriched with magnetically active isotopes like ^{1}H, ^{13}C, ^{15}N or ^{31}P, and held in a powerful magnetic field, can be explored at atomic resolution. The power of NMR lies in its ability to characterize the molecular environment of the nucleus and makes it pervasive throughout molecular, material and biological sciences as well as in medicine. NMR experiments can be efficiently designed to probe bond angles, distances between nuclei, coupling constants, chemical shifts, kinetic rate constants, etc.

The atomic-level understanding of biological process requires the determination of structural models of representative conformers of the biomolecules involved. The nature of the amino acid residues and their organization in space determine the character of a protein to either change its shape frequently or remains mostly rigid. It is well established that transitions between different states of a protein is crucial to perform important biological functions.[58] Proteins can switch to their conformation on many different timescales to attain biological functionality. The structural flexibility of proteins enables them to play essential role virtually in all cellular functions. The understanding of these dynamic processes requires the description of the ensemble of positions and time scales of fluctuations for all atoms in a protein undergoing structural fluctuation. A detailed, quantitative description of internal dynamics of proteins at atomic level aims *in fine* at understanding their biological function.

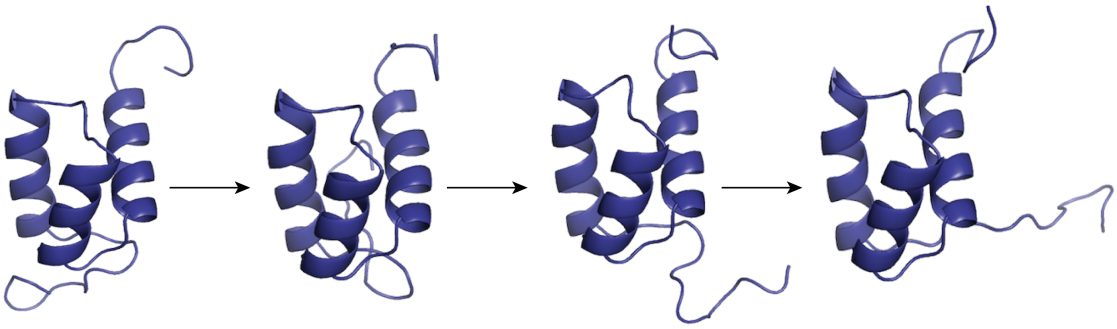


Figure 23: Proteins are not strictly static molecules, but their conformations fluctuate and interconvert in time. These fluctuations occur at different length and time scales and are linked to relevant functions. Different conformers of Engrailed-2 homeodomain in solution solved by NMR spectroscopy have been shown here (pdb code 3ZOB).[31]

Constant methodological developments over the last three decades empower NMR spectroscopy to probe molecular motions on a broad range of time scales in both ordered and disordered proteins. The structural flexibility of IDPs makes it necessary to adapt the concepts and methods to describe and understand their dynamic properties as compared to their folded counterparts.[59] Relaxation is the process by which non-equilibrium magnetization returns to the equilibrium state and is exquisitely sensitive to molecular motion (Figure 24) from fast (ps–ns) to slow (µs–ms) time scales.[60] NMR spectroscopy provides a plethora of different relaxation experiments to investigate the motional dynamics on different time scales.

The most common set of nuclear spin relaxation experiments that is used to probe the dynamics of the protein backbone is the measurement of nitrogen-15 longitudinal R_1, and transverse R_2 relaxation rates, accompanied by the measurement of the steady-state ^{15}N-{^1H} nuclear Overhauser effect (NOE). These rates provide information on the fluctuations of the ^{15}N-^1H dipole-dipole interactions and ^{15}N chemical shift anisotropy. The auto-correlation function of these interactions is not directly measured but its Fourier transform, the spectral density function is sampled at the transition frequencies of the spin system (e.g. the Larmor frequency of the proton and nitrogen-15 nuclei).[23] Mapping the spectral density function is the least invasive and most "model free" way to describe motions of bond vector, thus providing quantitative parameters for models of overall and local motions in folded proteins and the distribution of motions in disordered proteins.[24,25]

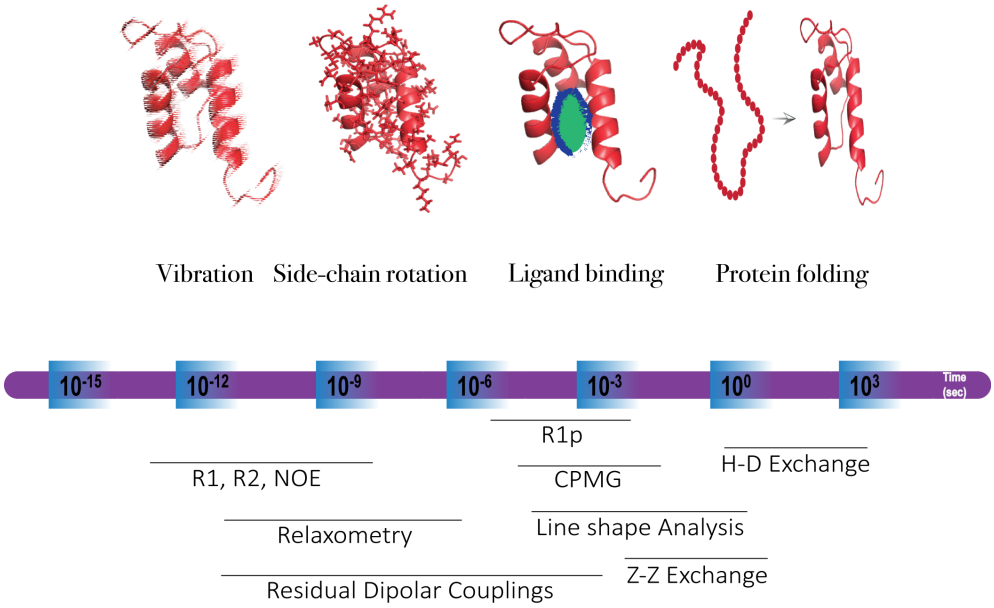


Figure 24: The different time scale of motions in proteins that can be efficiently characterized by NMR. The detection of motions on time scales from picoseconds to hours makes NMR a unique tool to understand relevant biological functions at atomic level.

The present chapter describes the new approach to characterize the timescale of motions in disordered region of Engrailed-2.

4.2 Nuclear Magnetic Relaxation and Its Relationship to Dynamics

NMR spin relaxation is a phenomenon in which the nuclear spin polarization perturbed away from the equilibrium state returns back to equilibrium thanks to random fluctuations of the local magnetic fields. The local magnetic fields are generated by the dipole-dipole, chemical shift anisotropy (CSA), quadrupolar coupling, isotropic chemical shift, and scalar coupling Hamiltonians.[61] The relaxation rate constants, corresponding to many spin terms (polarizations on various nuclei, in-phase or anti-phase, single- or multiple-quantum, etc.) can be measured experimentally. The measured relaxation rates reveal the time scales of the molecular motions that directly depend on the magnitude of the oscillating fields modulated by dynamic processes at the atomic level.[60]

The nitrogen-15 relaxation rates (^{15}N R_1, R_2 and ^{15}N-{^1H} NOE) are usually measured using two-dimensional (HSQC-type) experiments to provide information of backbone NH group dynamics in proteins.[62] In addition to these typically measured experiments, longitudinal (R_{ccl}) and transverse (R_{cct}) cross-relaxation rates, resulting from the cross relaxation of the ^{15}N CSA and the dipolar interaction of the ^{15}N-^1H, can be conveniently used to estimate the spectral density with no perturbation from chemical exchange and thus to quantify chemical exchange contribution to typically measured R_2 relaxation rates. Chemical exchange is a process in which a nucleus exchanges between two or more environments that contributes to chemical shift, scalar coupling, dipolar coupling and relaxation rate. [63,64]

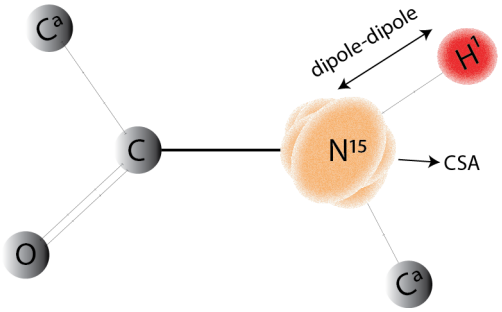


Figure 25: Principal interactions that contribute to ^{15}N relaxation. The relaxation results mostly from the fluctuations of the ^{15}N-^1H dipole-dipole interactions and ^{15}N chemical shift anisotropy.

Motions on ps-ns time scales can be characterized by amide ^{15}N relaxation resulting from the fluctuations of the ^{15}N-^{1}H dipole-dipole interaction and the ^{15}N chemical shift anisotropy. Motions on μs-ms time scales can be studied by spin relaxation resulting from modulation of isotropic chemical shifts.[60]

As its name indicates, the dipole-dipole interaction is the interaction between the ^{15}N nucleus and the dipolar magnetic field of the closest (i.e. directly bound) ^{1}H nucleus. The rotation of the internuclear ^{15}N-^{1}H vector due to molecular tumbling of the protein and local motions leads to fluctuations of the local dipolar field from ^{1}H nucleus at the position of the ^{15}N nucleus. This fluctuation of the dipole-dipole interaction contributes to relaxation.

The chemical shift interaction describes the local shielding of the static magnetic field by the electrons at the position of the nitrogen-15 nucleus. The isotropic part gives the chemical shift measured in high-resolution spectra. The anisotropic part depends on the orientation of the molecular orbitals and hence of the peptide plane with respect to the static magnetic field. Similarly, overall rotational diffusion of the protein and local motions, lead to fluctuations of this shielding, which contribute to relaxation.[61] Importantly, we can consider that the fluctuations of the CSA and dipole-dipole interactions are fully correlated and we can consider both fluctuations to account for motions of the internuclear ^{15}N-^{1}H vector.

The motions of the internuclear ^{15}N-^{1}H vector can be described by the time-dependent rotational correlation function C(t). The shape and decay of C(t) contain the information about the overall tumbling and the internal motions for each internuclear ^{15}N-^{1}H vector in the protein[23] (Figure 26).

Relaxation rates do not depend directly on the correlation function C(t) but on the values of its Fourier transform, the spectral density function J(ω), at the eigenfrequencies of the spin system.[62]

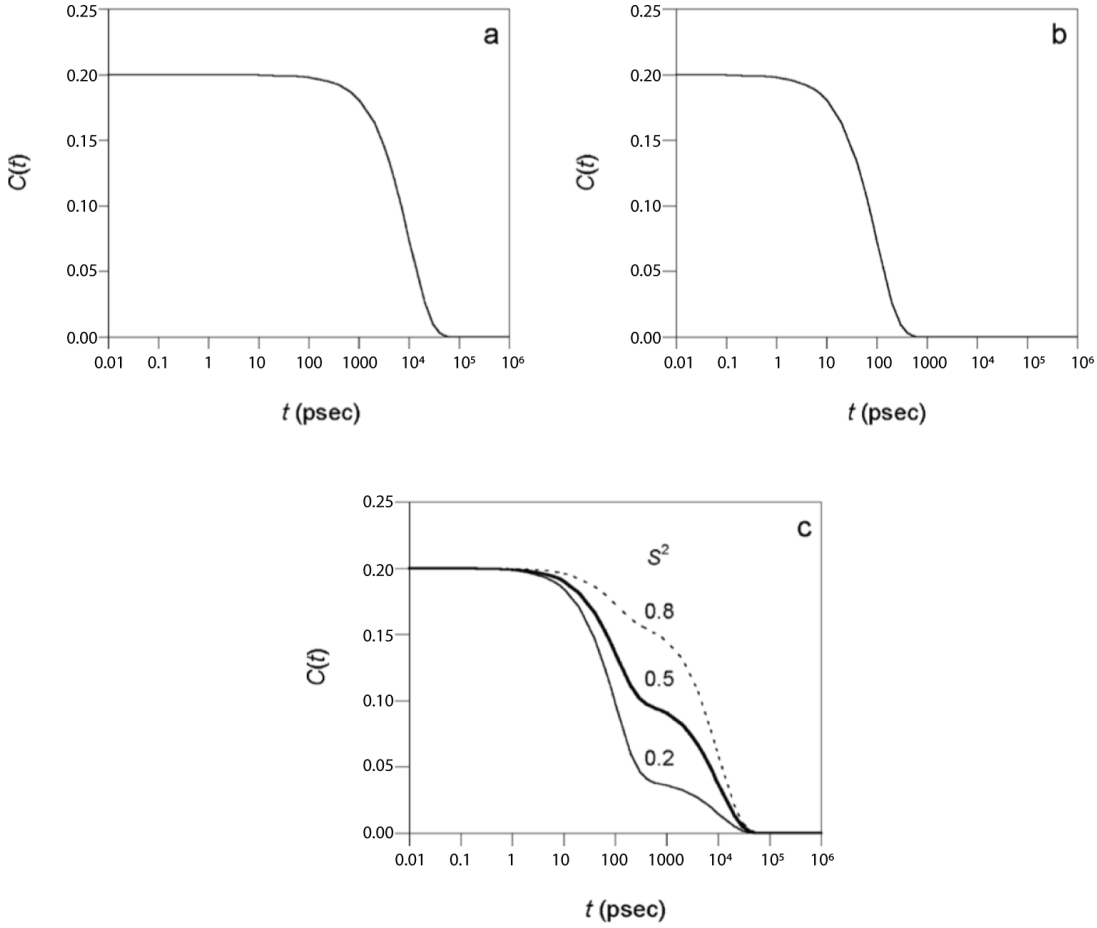


Figure 26: Graphical representation of the rotational correlation function C(t). (A) Isotropic tumbling with a single correlation time (τ_c = 10 ns). (B) Slow isotropic tumbling (τ_c = 10 ns) with hypothetical faster (τ_e = 100 ps) isotropic internal motions (C) Slow isotropic tumbling (τ_c = 10 ns) with faster (τ_e = 100 ps), spatially restricted isotropic internal motions with different order parameter S^2 values of 0.2, 0.5 and 0.8. Figures are adapted from Ref.62.[62]

4.3 Spectral density mapping

The aim of most dynamics studies is to characterize in details the motions of internuclear ^{15}N-^{1}H vectors. A first step can be to first quantify the spectral density function J(ω). The derivation of the spectral density function J(ω) at the eigenfrequencies of the spin system from relaxation rate constants is called spectral density mapping.[24] The commonly measured relaxation rates depend on the spectral density function as follows,

$$R_1 = R_1^{DD} + R_1^{CSA} \quad (1)$$

$$R_1^{DD} = \frac{1}{4} d^2 [J(\omega_H - \omega_X) + 3J(\omega_X) + 6J(\omega_H + \omega_X)]$$

$$R_1^{CSA} = c^2 J(\omega_X)$$

$$R_2 = R_2^{DD} + R_2^{CSA} + R_{ex} \quad (2)$$

$$R_2^{DD} = \frac{1}{8} d^2 [4J(0) + J(\omega_H - \omega_X) + 3J(\omega_X) + 6J(\omega_H) + 6J(\omega_H + \omega_X)]$$

$$R_2^{CSA} = \frac{1}{6} c^2 [4J(0) + 3J(\omega_X)]$$

$R_{ex} =$ conformational exchange broadening contribution

$$NOE = 1 + \frac{1}{4} d^2 T_1 \left(\frac{\gamma_H}{\gamma_X}\right) [6J(\omega_H + \omega_X) - J(\omega_H - \omega_{XH})] \quad (3)$$

$$\sigma_{XH} = \frac{1}{4} d^2 [6J(\omega_H + \omega_X) - J(\omega_H - \omega_X)] \quad (4)$$

where DD and CSA are the dipole-dipole and chemical shift anisotropy relaxation mechanisms, respectively. ω_H and ω_X are the angular Larmor frequencies for ^1H and nucleus X. c and d are the constants defined as $c = \omega\Delta\sigma/\sqrt{3}$ where $\Delta\sigma$ is the CSA of the X spin and $d = \left(\mu_0 h \frac{\gamma_X \gamma_H}{8\pi^2}\right)\{\frac{1}{r_{XH^3}}\}$ where μ_0 is the permeability of free space; h is Planck's constant and γ_H and γ_X are the gyromagnetic ratios of ^1H and X nuclei, respectively.[65]

The complete evaluation of spectral density function to characterize the motion of each NH bond vector requires the evaluation of the spectral density function sampled at the five frequencies, J(0), J(ω_N), J(ω_H), J($\omega_H + \omega_N$), J($\omega_H - \omega_N$).[62] The first practical implementation of the spectral density mapping at above-mentioned frequencies in eglin C has clearly demonstrated that J(ω_H), J($\omega_H + \omega_N$), J($\omega_H - \omega_N$) have very similar values but differ markedly to J(0) and J(ω_N) frequencies.[66] This observation led to suggest a particularly elegant and robust approach called "reduced spectral density mapping". Only, three nitrogen-15 relaxation experiments are necessary (R_1(^{15}N), R_2(^{15}N) and ^{15}N-{^1H} NOE), the difficult measurement of the relaxation rates of proton polarizations is avoided. The spectral density functions is determined at three frequencies only J(0), J(ω_N), J($0.87\omega_H$)[24,25], which provides most of the information of the internal dynamics by the relationship of relaxation parameters to, reduced spectral density values can be written as,

$$R_1 = \left(\frac{d^2}{4}\right)[3J(\omega_N) + 7J(\beta_1\omega_H)] + c^2J(\omega_N) \quad (5)$$

$$R_2 = \left(\frac{d^2}{8}\right)[4J(0) + 3J(\omega_N) + 13J(\beta_2\omega_H)] + \left(\frac{c^2}{6}\right)[4J(0) + 3J(\omega_N)] \quad (6)$$

$$\sigma_{NH} = \left(\frac{d^2}{4}\right)5J(\beta_3\omega_H) \quad (7)$$

in which $\beta_1 = 0.921$, $\beta_2 = 0.955$, and $\beta_3 = 0.87$. The system of equations can be solved for $J(0)$, $J(\omega_N)$, and $J(\beta_3\omega_N)$ in one of three ways: method (1), assuming a constant spectral density function at high frequency $J(\beta_i\omega_N) = J(\beta_3\omega_N)$; method (2) assuming a quadratic decay of the spectral density function $J(\beta_i\omega_N) = (\beta_3/\beta_i)^2 x J(\beta_3\omega_N)$; or method (3) evaluating $J(\beta_i\omega_N)$ from the static magnetic field dependence of $J(\beta_3\omega_N)$ under the assumption that the spectral density function is locally linear.[60] The latter method is only applicable if data at two fields or more are available.

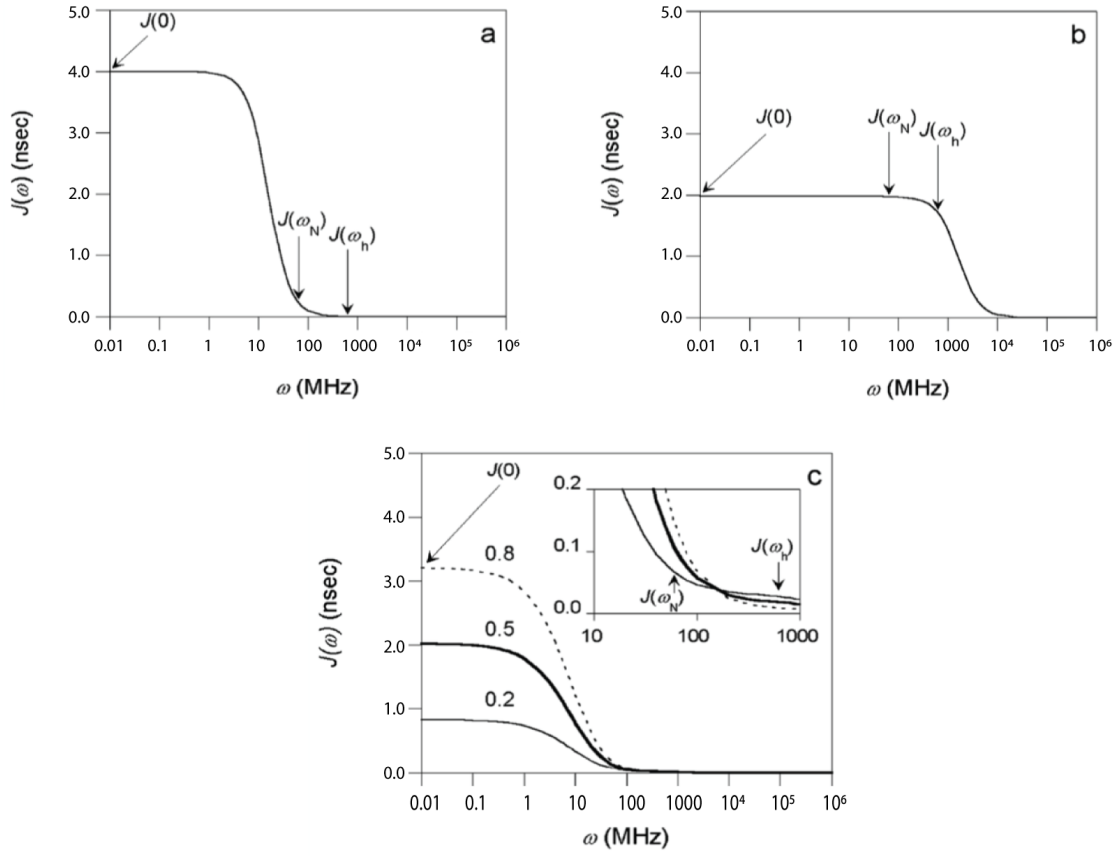


Figure 27: Graphical representations of the reduced spectral density function corresponding to the rotational correlation functions C(t) shown in Figure 27. Figures are adapted from Ref.62.[62]

Longitudinal and transverse CSA/DD cross-correlated relaxation rates η_z and η_{xy} can be efficiently used to derive the spectral density function $J(0)$ from $J(\omega_N)$ to avoid contributions from line broadening due to chemical exchange in R_2 rates.[67] The

spectral density values, obtained from reduced spectral density mapping, contain information about the correlation time that describes the reorientational motions of each internuclear ^{15}N-^{1}H vector.

4.4 Separation of local and overall motions in globular protein

The analysis of the spectral density function in folded proteins relies on a few approximations. The first approximation is the separation between overall rotational diffusion and internal motions, which allows for the factorization of the correlation function. This approximation has been challenged[68] but is most often justified by the fact that the correlation times for overall tumbling and internal motions are distinctly different. The popular "model-free" formalism introduced by Lipari and Szabo, offers a simple description of internal motions, described by an order parameter S^2 and effective correlation time for internal motions τ_{eff}.[69] For an isotropically reorienting macromolecule, the model-free spectral density is:

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_a}{1+(\omega\tau_a)^2} + \frac{(1-S^2)\prime\tau_b}{1+(\omega\prime\tau_b)^2}\right) \qquad (8)$$

with $\tau_b^{-1} = \tau_a^{-1} + \tau_{eff}^{-1}$, τ_a is the correlation time for the overall tumbling and τ_{eff} is the effective correlation time for local motions. S^2 is the model-free order parameter.

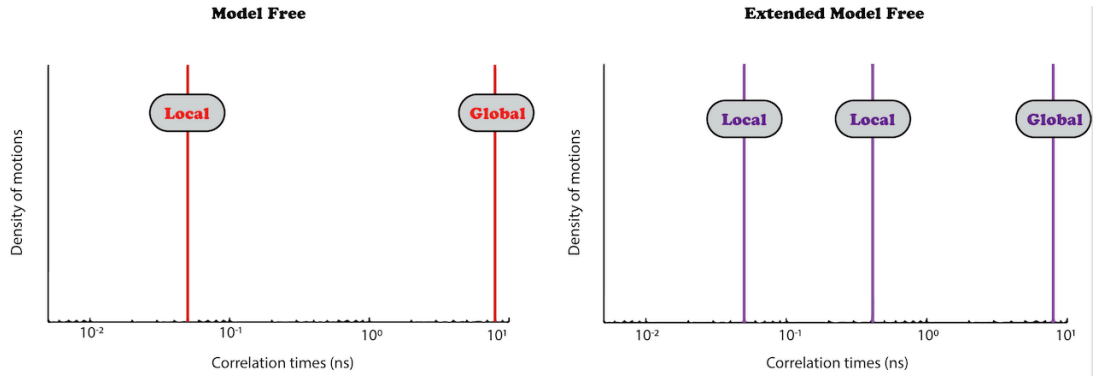


Figure 28: Model free (2CT) and Extended Model free (3CT). In the equivalent 2CT and 3CT analysis in IDPs, both the value and the relative weight of each correlation time are adjusted (see below).

The extended model free (EMF) introduces a second correlation time for internal motions. This is necessary when internal motions of significant amplitude occur on both fast and slow time scales. EMF introduces two correlation times τ_f, τ_s and respective order parameters S_f, S_s.[70] For macromolecules reorienting isotropically, the "extended model-free" spectral density is:

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_a}{1+(\omega\tau_a)^2} + \frac{(S_f^2-S^2)\tau_b}{1+(\omega\tau_b)^2} + \frac{(1-S_f^2)\tau_c}{1+(\omega\tau_c)^2}\right) \quad (9)$$

with $\tau_c^{-1} = \tau_a^{-1} + \tau_f^{-1}$; $\tau_a > \tau_b > \tau_c$; S_f^2 is equivalent to the extended model-free order parameter for the fast process with a correlation time τ_f.

4.5 Limits of the application of model-free approaches to IDPs

There is not yet a specific model to describe reorientational motions in the case of IDPs. The broad conformational space of IDPs prevents the definition of an overall diffusion tensor. Even if semi-local and local motions could be defined, the statistical dependence of these two types of motions would not be certain. The Model-Independent Correlation time distribution (MIC) approach[71] uses a formalism mathematically identical to the extended model free approach. The results are then used to extract parameters that describe the distribution of correlation times. This approach is legitimate as long as the distribution of correlation times has a single mode. This is not necessarily expected given the complexity of conformational ensembles of IDPs.

Several approaches have been developed in the last 15 years to extract quantitative information about pico- and nanosecond dynamics in IDPs and IDRs from nuclear spin relaxation rates. Most of these approaches are based on the model-free formalism, with the use of residue-specific correlation times[72] (i.e., no overall diffusion tensor), the introduction of a distribution of fast picosecond and slower nanosecond motions[73], or a statistical analysis of extended model-free parameters.[71]

The analysis with distributions of correlation times necessarily introduces a physical bias through the choice of a given mathematical framework to describe the

distribution of correlation times. Importantly, both the MIC and distribution approaches seem inappropriate to describe accurately the distribution of correlation times, which is a priori unknown. Moreover, simply increasing the number of correlation times or distributions in either approach would be challenging since the empirical information available from nuclear spin relaxation is limited.

4.6 Interpretation of Motions by a Projection onto an Array of Correlation Times (IMPACT)

We introduce a novel approach for the Interpretation of Motions by a Projection onto an Array of Correlation Times (IMPACT) to analyze multiple-field relaxation data in disordered proteins. This method is designed to rely as little as possible on any particular physical model of protein motions. We define an array of correlation times τ_i (with reciprocal frequencies $\omega_i = 1/\tau_i$) in a range that is effectively sampled by ^{15}N relaxation. The experimental spectral density function is then reproduced by a sum of Lorentzian functions $J_i(\omega)$, one for each correlation time τ_i. The result of this process, similar to a projection onto a basis of Lorentzian functions, is a discrete distribution of correlation times in a range relevant for relaxation. The multimodal character of the distribution of correlation times can thus be revealed, and the most relevant correlation times for backbone motions clearly emerge.

4.6.1 Principle and Optimization of IMPACT

The limitations of conventional approaches to the analysis of relaxation rates in IDPs and IDRs lies in the complexity of their dynamics, which span at least three orders of magnitude. Thus, it seems unlikely to characterize accurately the relaxation rates by a single distribution of correlation times, or by a small number of correlation times. However, the scarcity of relaxation rates limits the number of adjustable parameters that can be determined, and thus limits the sophistication of spectral density functions that can be postulated.

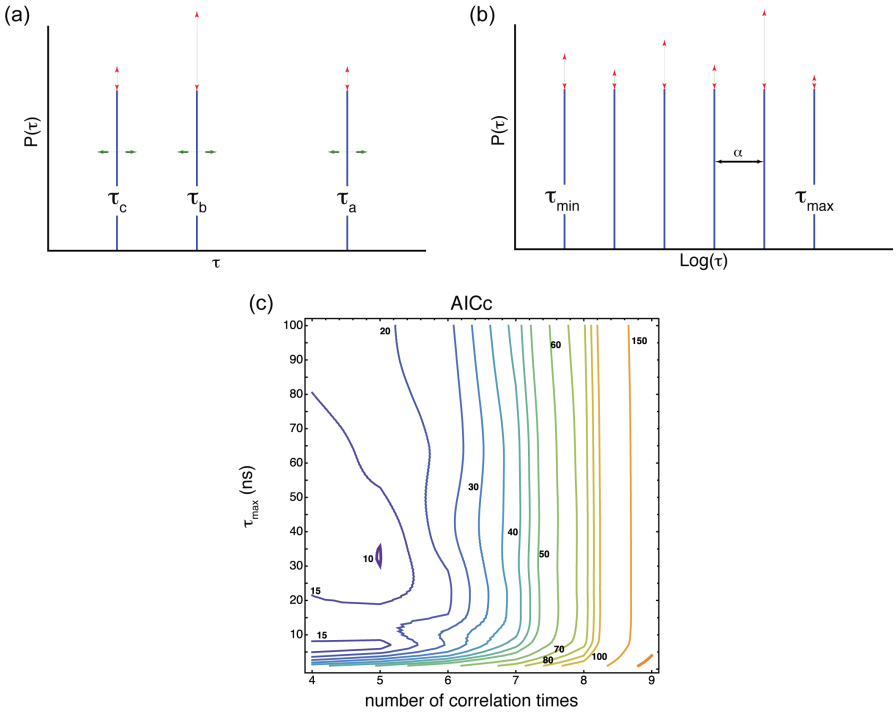


Figure 29: Principle and optimization of the parameters of the IMPACT analysis. (a) In the 3CT analysis, both the value and the relative weight of each correlation time must be adjusted. (b) In IMPACT, the values of the correlation times are fixed and equally spaced on a logarithmic scale, so that only their relative weights need adjustment. (c) Optimization of IMPACT. The AICc is shown. The range $[\tau_{min}, \tau_{max}]$ of time scales characterized by IMPACT was varied from [1 ps, 1 ns] to [100 ps, 100 ns] and the number of correlation times was varied from n = 4 to 9.

Herein, we significantly increase the number of correlation times by defining an array of fixed correlation times. Only the relative coefficient of each correlation time in the distribution is fitted to experimental data, so that the number of adjustable parameters is reduced. Thus, the only assumption is that the correlation function can be approximated by a sum of exponentials. The array of correlation times is defined as a geometric series, so that correlation times are equally spaced on a logarithmic scale (Figure 29 a-b):

$$\tau_i = \alpha^{i-1}\tau_{max} \qquad \alpha = (\tau_{min}/\tau_{max})^{\frac{1}{n-1}} \tag{10}$$

Thus $J(\omega)$ is a sum of Lorentzian functions:

$$J(\omega) = \sum_{i=1}^{n} J_i(\omega) = \frac{2}{5}\sum_{i=1}^{n} \frac{A_i \tau_i}{1+(\omega\tau_i)^2} \quad (11)$$

where A_i is the coefficient of the correlation time τ_i in the spectral density function.

The coefficients A_i must fulfill the following normalization constraint:

$$\sum_{i=1}^{n} A_i = 1 \quad (12)$$

Thus, the number of free parameters is $n - 1$.

A preliminary step of the IMPACT analysis is the optimization of the three parameters τ_{min}, τ_{max} and n. Although the correlation times could be chosen as the inverse of the Larmor frequencies at which the spectral density is mapped,[74] we have adopted a different approach. We have first defined the range of correlation times that were sampled by various ^{15}N relaxation rates. We considered that the lowest magnetic field adapted to protein studies was 400 MHz, while the highest accessible field currently was 1 GHz for our study (discussed below). Thus, the lowest non-zero frequency at which the spectral density was sampled is $\omega_N/2\pi$ = 40 MHz, and the highest $0.87\omega_H/2\pi$ = 870 MHz. A Lorentzian function with a correlation time τ_c = 40 ns drops to 1% of *J(0)* at $\omega_N/2\pi$ = 40 MHz. A Lorentzian with τ_c = 18 ps merely decreases to 99% of J(0) at $0.87\omega_H/2\pi$ = 870 MHz. The resulting range spans slightly more than three orders of magnitude. Therefore, we have decided to limit the range to three orders of magnitude

$$\tau_{max}/\tau_{min} = 10^3 \quad (13)$$

In order to define the optimal values of τ_{min} and τ_{max}, we have carried out a series of IMPACT analyses for [τ_{min}, τ_{max}] = [1 ps, 1 ns] to [100 ps, 100 ns] as well as for $4 < n < 9$. The statistical relevance of each combination of parameters was evaluated from the resulting second order variant of AIC (Akaike's Information Criteria) also called AICc.[75]

$$AICc = AIC + \frac{2n_{model}(n_{model}+1)}{n_{exp}-n_{model}-1} \quad (14)$$

with $$AIC = n_{exp} ln\left(\frac{\chi^2}{n_{exp}}\right) + 2n_{model} + C$$

where n_{exp} is the number of experimental data and $n_{model} = n$-1 is the number of free parameters in each model. C is a constant.

4.7 Results and discussion

In order to characterize the time scale of motions in disordered region of Engrailed-2, ^{15}N relaxation rates were measured at multiple fields for the implementation of IPMACT approach.

4.7.1 Relaxation experiments

Relaxation experiments were carried out at 400, 500, 600, 800, and 1000 MHz (corresponding to proton Larmor frequencies) to determine longitudinal R_1 ^{15}N relaxation rates, the steady-state ^{15}N-{^{1}H} nuclear overhauser effect (NOE), as well as the longitudinal η_z and transverse η_{xy} cross-relaxation rates due to correlated fluctuations of the ^{15}N chemical shift anisotropy (CSA) and the dipolar coupling with the amide proton (Figure 30).

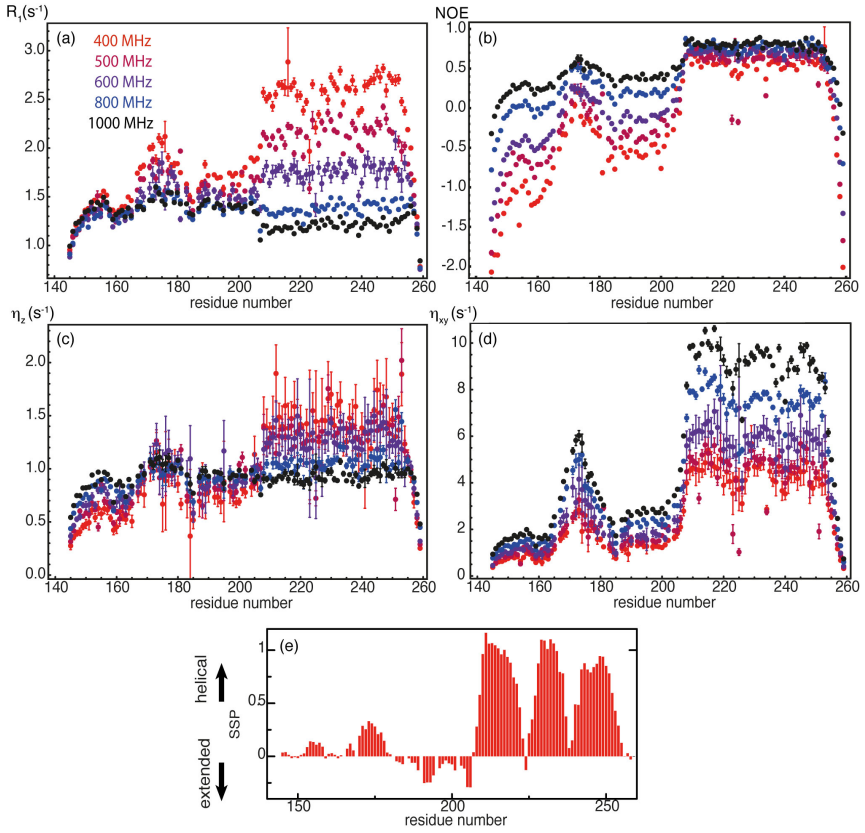


Figure 30: Backbone ^{15}N relaxation rates and nuclear Overhauser effects measured in Engrailed at five magnetic fields: 400 MHz (red), 500 MHz (burgundy), 600 MHz (purple), 800 MHz (blue) and 1000 MHz (black). (a) Longitudinal relaxation rates R_1 of ^{15}N; (b) ^{15}N-{^{1}H} NOE ratios; (c) longitudinal cross-relaxation rates η_z due to correlated fluctuations of the ^{15}N CSA and the ^{15}N-^{1}H dipolar couplings; (d) transverse cross-relaxation rates η_{xy} due to the same correlated fluctuations. (e) Secondary structure propensity (SSP) calculated from the chemical shifts of carbonyl, alpha and beta carbon-13 nuclei.[56]

The uniform decrease of the longitudinal relaxation rates R_1 with increasing magnetic field B_0 in the homeodomain (Figure 30a) indicates motions in the nanosecond range, resulting from overall rotational diffusion. The variations $R_1(B_0)$ are much less pronounced in the disordered region, except in the 169-174 hexapeptide. Longitudinal cross-relaxation rates η_z (Figure 30c) increase with B_0 in the IDR, which reflects the very slow decay (slower than $1/\omega$) of the spectral density function in the range 40-100 MHz (i.e., the range of Larmor frequencies ω_0(^{15}N) between 9.4 and 23.5 T). Nuclear Overhauser effects are sensitive markers of local order in IDPs and IDRs and have been used as such for many years.[76] Indeed, the variations of NOEs along the sequence are pronounced at moderate fields (9.4 to 14.1 T), however, the profile of NOEs is much flatter at high fields, in particular at 23.5 T. On the other hand,

transverse cross-correlation rates η_{xy}, which depend primarily on $J(\omega = 0)$, exhibit sharp variations at all fields and are strongly correlated with SSP scores. The profile of transverse relaxation rates R_2 is marked by variations of both order and line broadening due to chemical exchange (Figure 31).

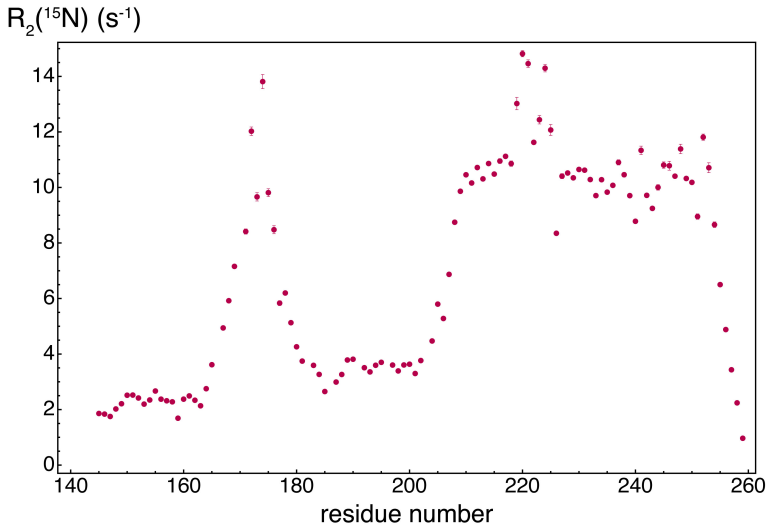


Figure 31: ^{15}N transverse relaxation rates R_2, measured at 18.8 T under a Carr-Purcell-Meiboom-Gill (CPMG) train of 180º pulses.

4.7.2 Spectral density mapping

Most current software packages designed to characterize protein dynamics based on relaxation rates offer a direct derivation of the parameters of local dynamics (order parameters and correlation times for local motions). This approach is efficient and reliable when the theoretical framework of the analysis has been validated. Since a general understanding of motions in intrinsically disordered proteins is still lacking, the derivation of the spectral densities from relaxation rates provides a representation of experimental data that is more amenable to physical analysis than relaxation rates.[77] Spectral density mapping can be achieved without resorting to any proton auto-relaxation rate.[78] The effective spectral density at high-frequency $J(0.87\omega_H)$ (Figure 32b) can be derived from {^{1}H}-^{15}N NOE and longitudinal ^{15}N relaxation rates for different magnetic fields according to,

$$J(0.87\omega_H) = \frac{4\gamma_N R_1 (NOE-1)}{5d^2\gamma_H} \quad (15)$$

with $$d = \frac{\mu_0}{4\pi}\frac{\hbar\gamma_H\gamma_N}{\langle r_{NH}^3\rangle}$$

μ_0 is the permittivity of free space; γ_H (γ_N) is the gyromagnetic ratio of the proton (respectively of the ^{15}N nucleus); $\hbar$ is Planck's constant divided by 2π and r_{NH} = 1.02 Å is the distance between the amide proton and the ^{15}N nucleus.

Our data recorded at five magnetic fields allowed us to fit the spectral density function at high frequency $J(0.87\omega_H)$ to the following expression,

$$J(\omega) = \lambda + \frac{\mu}{\omega^k} \quad (16)$$

as in an earlier study of ^{13}C relaxation.[79] The parameters λ and μ are real positive numbers, while k = 1 or 2. Representative results of fits to (16) are shown in Figure 32d-f. A good fit has been obtained with a quadratic term (k = 2) for 37 out of the 59 analyzed residues of the homeodomain. Note that more than half of the residues that can be fitted with a function with k = 1 of the homeodomain belongs to mobile regions, in particular to its N- and C-termini and to the loop α_1-α_2. In the disordered region of the protein, the best fits are obtained with k = 1 for most residues, *i.e.*, for 39 out of 49 residues (between residues 146 and 199). Moreover, 5 of the ten residues that can be best fitted with k = 2 belong to the ordered region around the hexapeptide. The prevalence of k = 1 is likely a consequence of a broader distribution of the true correlation times of reorientational motions.

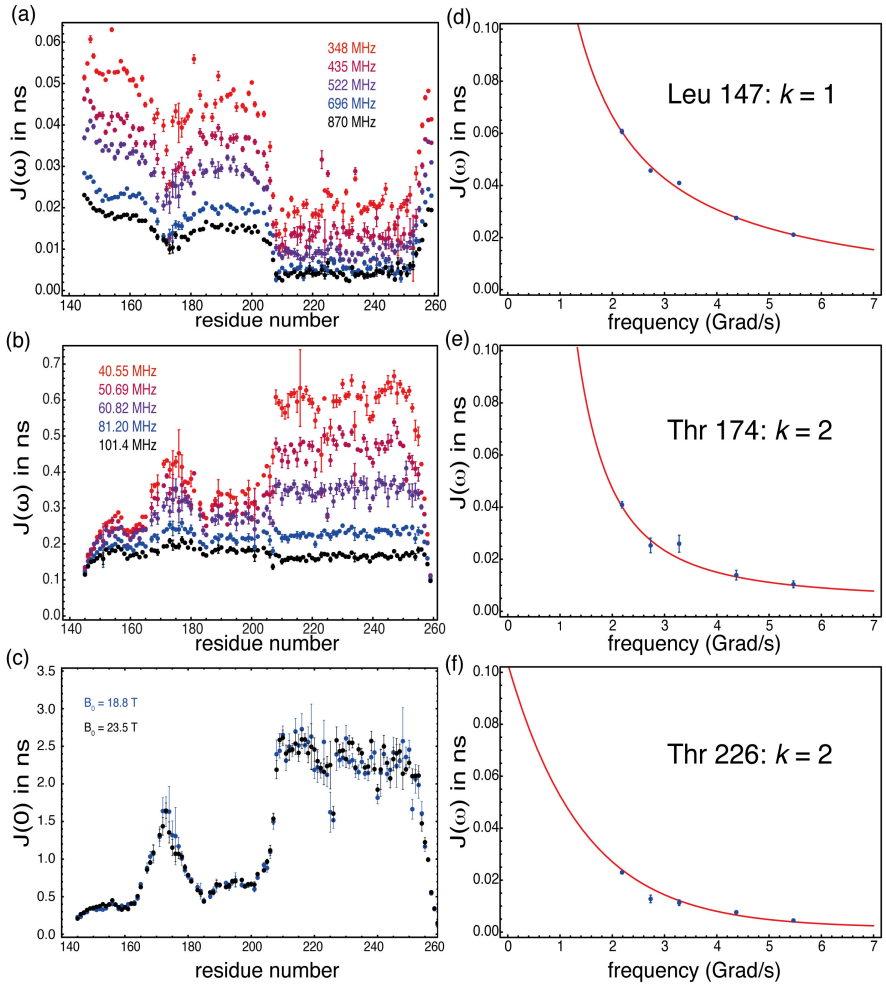


Figure 32: Spectral density functions for backbone NH vectors in Engrailed 2. (a) Effective spectral density near the proton Larmor frequency $J(0.87\omega_H)$ (ns). (b) Spectral density at the Larmor frequency of nitrogen-15, $J(\omega_N)$ (ns). (c) Spectral density at zero frequency, $J(0)$ (ns). All data are color-coded as a function of the magnetic field at which the relaxation data were recorded. The spectral density function $J(0.87\omega_H)$ in ns as a function of frequency in Grad/s (10^9 rad/s) is shown as a function of frequency for three selected residues: (d) Leu147, (e) Thr174, and (f) Thr226, along with the best fit to the function $\lambda + \mu/\omega^k$ with k = 1 or 2.

The results of the fit of the spectral density were used to evaluate the contribution of the spectral density at high frequency (at $\omega_H \pm \omega_N$) in the derivation of $J(\omega_N)$ from the rate R_1 according to the equation,

$$J(\omega_N) = \left(\frac{R_1}{(3d^2/4+c^2)} - \frac{(6J(\omega_H+\omega_N)+J(\omega_H-\omega_N))}{(3+4c^2/d^2)}\right) \qquad (17)$$

with $c = \gamma_N B_0 \Delta\sigma/\sqrt{3}$. B_0 is the magnetic field and $\Delta\sigma$ is the axially symmetric chemical shift anisotropy of the ^{15}N nucleus ($\Delta\sigma$ = 160 ppm).

Overall, contributions of high-frequency terms to R_1 are small,[80] so that the deviations between the values of $J(\omega_N)$ obtained from a series of approximations[66] and the current method are limited to about 2%. The values of $J(\omega_N)$ derived at five magnetic fields are shown in Figure 32b.

To avoid contributions from line-broadening due to chemical exchange, we did not consider transverse relaxation rates $R_2(^{15}N)$ and used longitudinal and transverse CSA/DD cross-correlated relaxation rates η_z and η_{xy} to derive the spectral density function $J(0)$ from $J(\omega_N)$ using,[67]

$$J(0) = J(\omega_N)\frac{3}{4}\left(2\frac{\eta_{xy}}{\eta_z} - 1\right) \quad (18)$$

The measurements of η_z and η_{xy} were not precise enough at lower fields to provide reliable estimation of $J(0)$, due to a lack of sensitivity (in particular for η_z). However, the data recorded at 18.8 T and 23.5 T were very similar and do not show any of the outliers observed at lower fields. The contributions of chemical exchange preclude the proper derivation of $J(0)$ from $R_2(^{15}N)$ rates, in particular at high magnetic fields.

4.7.3 Analysis with two- or three-correlation times

We first tried to fit two simple models, where the spectral density function is assumed to consist of a sum of two and three Lorentzians to describe motions with two or three correlation times, respectively. These models are identical to model-free and extended model-free approaches (equation (8) and (9)).

The two functions were fitted to the experimental spectral density, and a simple model selection step consisted in comparing the second order variant of AIC (see Equation 14), with n_{exp} = 11 and n_{model} = 3 for the 2CT analysis and n_{model} = 5 for the 3CT analysis.

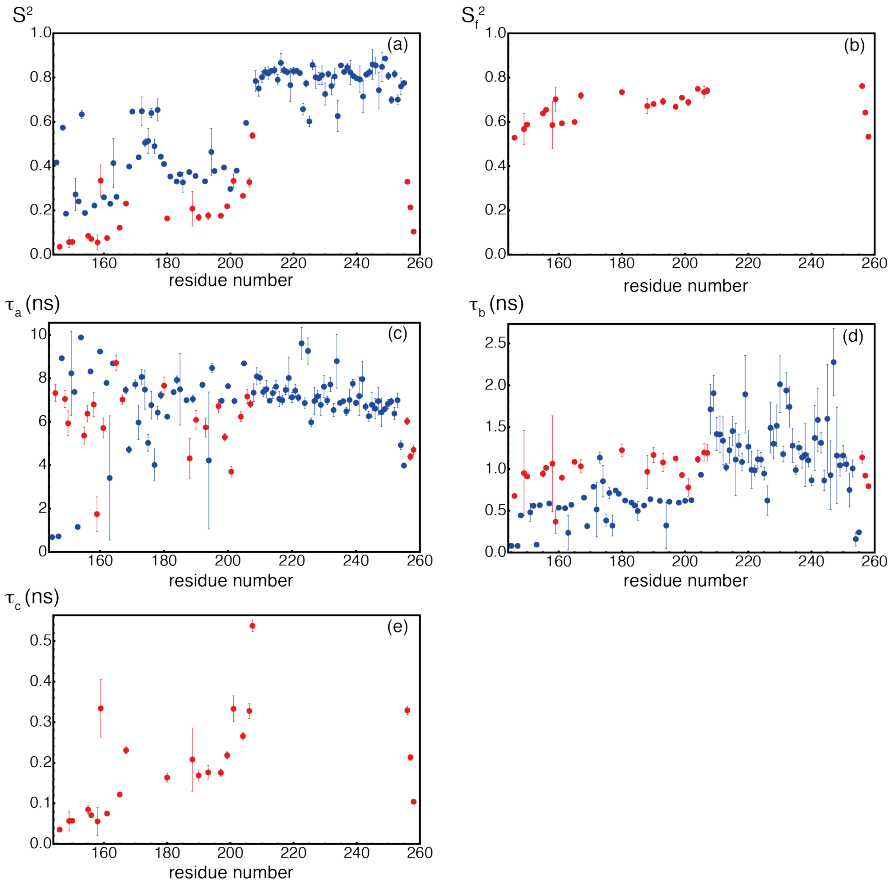


Figure 33: Results obtained for a conventional analysis with two (blue) or three (red) correlation times. (a) Order parameter S^2; (b) order parameter S^2_f for the faster motion in the 3CT analysis; (c) longest correlation time τ_a ; (d) intermediate correlation time τ_b ; and (e) shortest correlation time τ_c . The model (2CT in blue or 3CT in red) was selected based on the lowest AIC score.

Results of this analysis are shown in Figure 33. The 2CT analysis is clearly most suitable in the homeodomain (residues 200-260) and particularly in regions with a well-defined secondary structure. With few exceptions, the longest correlation time is a reliable measure of overall rotational diffusion. Interestingly, the second correlation time τ_b found for residues in the homeodomain lies in the range $1 < \tau_b < 2$ ns, with few exceptions. This could reflect widespread internal motions in the homeodomain. However, it is more likely due to fluctuations of the overall diffusion tensor, resulting from conformational transitions in the disordered N-terminal region on timescales between 1 and 100 ns.[81] In the disordered region (residues 146-199), the 3CT analysis is generally favored. Low order parameters S^2 are found throughout this region, with a significant increase in order in the hydrophobic cluster of the hexapeptide (residues 169-174). Surprisingly, the correlation times in the disordered

regions are fairly uniform, with a long correlation time $5 < \tau_a < 7$ ns, similarly to what was found in an unfolded protein[82] and very close to the overall tumbling of the homeodomain. The intermediate correlation time lies in the range $0.4 < \tau_b < 1.3$ ns, while the short correlation time for fast motions is $50 \text{ ps} < \tau_c < 300$ ps. The order parameter S^2 apparently absorbs most of the site-to-site variations of the spectral density function.

One should be careful with the physical interpretation of these results in the IDR of Engrailed. The three correlation times obtained are clearly spaced in time domain, which indicates a broad range of dynamic processes. The results should not be considered as actual correlation times of particular motions but as the best rendition of experimental results using two or three correlation times. This is illustrated by the jumps of order parameters and correlation times observed between 2CT and 3CT results in the IDR. For instance, it is difficult to assign the longest correlation time to a particular dynamical process in the absence of complementary experimental or computational information. Such a process could be a single well-defined type of motion, such as the rotational diffusion of an IDR segment. Alternatively, the longest correlation time might account for the tail of a distribution of correlation times and reflects slower motions in parts of the conformational space of the IDR. The presence of a broad distribution of correlation times can lead to a smooth decay of the spectral density function

4.7.4 Application of IMPACT to Engrailed

The optimal parameters $n = 5$ and $[\tau_{min}, \tau_{max}] = [32 \text{ ps}, 32 \text{ ns}]$ were employed to analyze the spectral density function in Engrailed.

Figure 34 illustrates the remarkable variety of dynamics found in Engrailed. In the homeodomain, the second correlation time τ_2 lies just below the correlation time for overall rotational diffusion, which is close to 7 ns, as can be seen from the analysis based on only two correlation times. Thus, the second coefficient is, by far, the most important in the homeodomain.

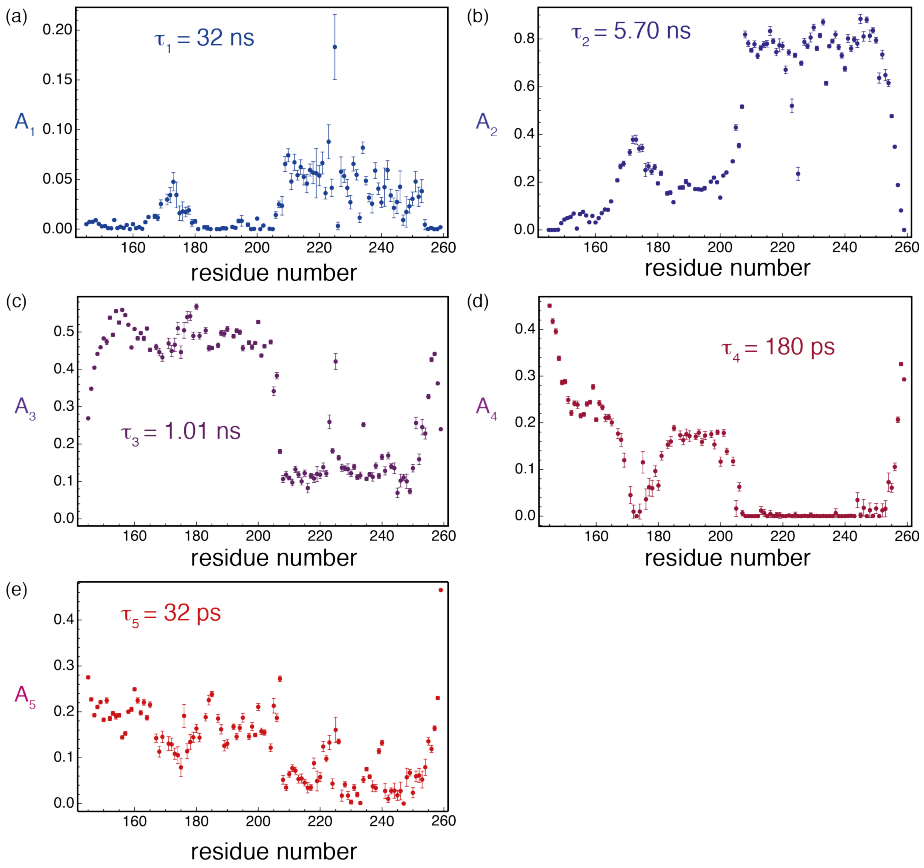


Figure 34: (a-f) Plot of the coefficients A_i for i = 1, 2... 5 of the n = 5 correlation times τ_i in the range $[\tau_{min}, \tau_{max}]$ = [32 ps, 32 ns] determined by the IMPACT analysis of Engrailed.

A small amplitude A_1 of τ_1 corrects for the fact that τ_2 is shorter than the actual correlation time for the motion of the whole domain. Small but significant and mostly uniform contributions A_3 of the correlation time τ_3 in the alpha helices may be attributed to fluctuations of the overall diffusion tensor, likely due to conformational fluctuations of the IDR (residues 146-199). Enhanced values of A_3 in the loops may reflect the flexibility of these regions.[31] The very small coefficients A_4 demonstrate the presence of a gap in the distribution of correlation times, as was also observed in ubiquitin.[83] Finally, the coefficients A_5 for the shortest correlation time τ_5 indicate the presence of fast motions in the tens of ps range.

Results obtained in the disordered region of Engrailed will be discussed with the help of Figure 34, but also with the 'IMPACT barcode' shown in Figure 35. The width of each histogram represents the coefficient A_i associated with the correlation time τ_1 that can be read on the vertical axis. This graph appears as a convenient way to

display in a single figure the results of the IMPACT analysis, as well as the results of the 3CT analysis (Figure 35a). For the sake of comparison, we define the coefficients $B_{a,b,c}$ associated with the correlation times $\tau_{a,b,c}$ as,

$$B_a = S^2, B_b = S_f^2 - S^2, B_c = 1 - S_f^2$$

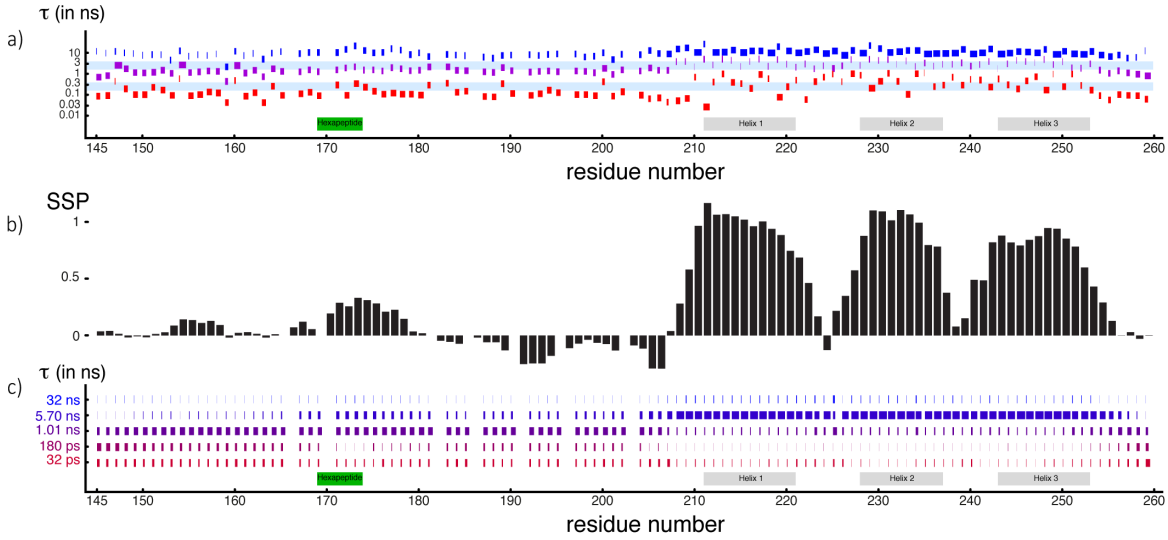


Figure 35: Graphical representations of (a) 3CT and (c) IMPACT analyses of the spectral density function in Engrailed. Histograms are drawn for all residues with the following rules: for each of the three correlation times $\tau_{a,b,c}$ determined in (a) and for each of the six correlation times τ_i (i = 1, 2, ...6) considered in (c), a rectangle represents the dynamics of the corresponding amino-acid residue along the horizontal axis. The width of each rectangle is proportional to the corresponding weights $B_{a,b,c}$ (a) or A_i (c). In (a), light blue horizontal bars represent the ranges of correlation times τ for which reciprocal frequencies lie between 40 <$1/(2\pi\tau)$ < 100 MHz or between 348 <$1/(2\pi\tau)$ < 870 MHz. Grey rectangles in (a) and (c) indicate alpha helices, and a green rectangle shows the location of the hydrophobic hexapeptide sequence. (b) The secondary structure propensity (SSP) is shown to guide the comparison between structural and dynamic features.

For the first residues at the N-terminus and the last residues at the C-terminus, significant coefficients $A_{3\text{-}5}$ are found for the three shortest correlation times. This seems to indicate the presence of motions that are broadly distributed over all time scales up to 1 ns. On the other hand, the two disordered regions just at the N-terminus and the C-terminus of the hexapeptide display a high density of motions around τ_3, in agreement with the 3CT analysis. The coefficients for the correlation time τ_4 decrease almost linearly with the distance to the N- or C-termini of the polypeptide chain in disordered regions. A notable difference between the disordered

region at the N-terminus and the one in between the hexapeptide and the homeodomain is the slight decrease of the coefficient A_3 and a significant increase of A_2. It is difficult to assign this change to a particular process without a better characterization of the conformational space of the protein. Nevertheless, two effects may contribute to the presence of some orientational order beyond 1 ns: first, this IDR is short and located between a folded domain and a small hydrophobic cluster, thus its dynamics are likely influenced by the overall diffusion of both structured elements; second, this IDR contains a majority of residues that favor extended conformations, as confirmed by SSP scores: 3 proline residues, 9 positively charged, and only 1 negatively charged residues between positions 177 and 198; which should restrict the conformational space and possibly slow down reorientational dynamics.

The barcode representation of IMPACT coefficients perfectly illustrates variations of the ensemble of correlation times between successive structural elements: in the hexapeptide, for instance, the decrease of motions in the sub-ns range is accompanied by an increase in the supra-ns range. Similarly the variation of the coefficients for correlation times at the N- and C-termini of the homeodomain (residues 200-210 and 254-260) illustrates the smooth transition of motional properties along the sequence of the polypeptide. Finally, even in the homeodomain, where a classical analysis of relaxation data should be most appropriate, the dynamic transitions between helices and the two loops are clearly visible and quantitatively characterized by the IMPACT approach. Loop α_1-α_2 features enhanced dynamics in both the 1 ns and tens of ps ranges, while loop α_2-α_3 shows a significant but more moderate enhancement of motions. Overall, the IMPACT representation offers an elegant view of the correlation of structural and dynamic features, as can be seen from the SSP scores.

The absence of a model selection step in IMPACT provides results that are directly comparable, residue by residue, which allows for a better qualitative description of the evolution of dynamics properties along the protein sequence. Such a description is more difficult in the 2CT and 3CT analysis, as can be seen from discontinuities in the transitions between residues analyzed with two or three correlation times in Figure 33.

4.8 Conclusions

We have presented a set of ^{15}N relaxation rates recorded at five magnetic fields on the 115-residue partially disordered protein Engrailed 2. The transverse cross-correlated rate η_{xy} was shown to be the most sensitive to the extent of order and disorder at all magnetic fields. The analysis validates the reduced spectral density mapping approach developed in the context of folded proteins. The spectral density functions can be fitted reasonably well with two or three correlation times but the results may be difficult to be interpreted. We have introduced a novel approach to the analysis of spectral density functions called 'interpretation of motions with a projection onto an array of correlation times' (IMPACT). This provides a better quantitative description of the spectral density function in IDRs, as found in Engrailed, than an analysis with three correlation times with the same number of adjustable parameters. We also introduce a "barcode" representation of IMPACT, which provides a condensed representation of a large amount of data in a single figure. This representation lends itself to qualitative discussions of the results (This work has been submitted).

4.9 Materials and Methods

4.9.1 Sample description

All experiments were performed on a sample of uniformly ^{15}N labeled Engrailed [residues 146-260] at a concentration of 0.6mM in 40mM succinate buffer at pH = 6 supplemented with 1 μg/mL of each of the three protease inhibitors Leupeptin, Pepstatin, AEBSF as well as 10μM EDTA, which allow one to increase the lifetime of the protein. The protein was prepared as described above in Engrailed-2 expression and purification protocol (chapter 3). Note that the sequence of the protein construct comprises three residues: Gly-Pro-Met at the N-terminus, before residue Glu146 left after cleavage of the GST-tag by PreScission protease. All experiments were carried out at 303 K, which was adjusted in each spectrometer to have a chemical shift difference between the signals of the methyl and hydroxyl protons in pure methanol of 1.462 ppm (4% protonated and 96% deuterated).

4.9.2 NMR spectroscopy

The relaxation rates have been measured at five different static fields of 9.4 T (proton frequency of 400 MHz), 11.7 T (500 MHz), 14.1 T (600 MHz), 18.8 T (800 MHz), and 23.5 T (1000 MHz). Three aliquots of the same sample were used for all experiments except at 11.7 T, which was performed on a separate, but identical, sample. For each field a full set of ^{15}N relaxation measurements was obtained. The longitudinal relaxation rates $R_1(^{15}\mathrm{N})$ were obtained in the traditional way,[84] with saturation of the water signal for each scan, while the transverse relaxation rates $R_2(^{15}\mathrm{N})$ were recorded with a train of ^{15}N π-pulses (CPMG pulse train), interleaved with ^{1}H π-pulses to suppress cross-correlated relaxation effects. ^{15}N-{^{1}H} nuclear Overhauser effects (NOE) were obtained by detecting the ^{15}N steady state polarization while saturating the protons with a train of π-pulses, with suitable inter-pulse delays and *rf* amplitudes.[85] Finally, experiments to measure the transverse and longitudinal cross-relaxation rates due to correlated fluctuations of the ^{15}N CSA with the dipolar coupling between the ^{15}N nucleus and the amide proton were recorded according to

the so-called symmetrical reconversion principle.[63] All experiments were recorded on Bruker Avance spectrometers. Experiments at 500 MHz, 800 MHz, 1 GHz and the NOE at 600 MHz have been recorded on triple resonance indirect detection cryogenic probes equipped with z-axis pulsed field gradients. Experiments at 600 MHz were recorded on an indirect detection triple resonance probe with triple axis gradients with detector coils at room temperature. Experiments at 400 MHz were recorded on a liquid-nitrogen cooled cryogenic probe (Prodigy BBO) equipped with z-axis gradients.

4.9.3 Spectral density analysis

The full analysis was carried out with eleven points obtained with the reduced spectral mapping, $J(0.87\omega H)$ and $J(\omega N)$ at 5 fields and $J(\omega=0)$ calculated from relaxation rates measured at 23.5 T. Analyses with two sets of three magnetic fields used seven point on the spectral density function; $J(\omega=0)$ was derived from the relaxation rates measured at the highest magnetic field, *i.e.,* 18.8 T or 23.5 T. A Monte Carlo simulation with 510 steps was performed in order to evaluate the error of each parameter *Ai*. All simulations were carried out with *Mathematica* (Wolfram research).

Chapter 5 Long-range contacts of disordered region of Engrailed-2 and their contribution to DNA binding

Chapter 5 Long-range contacts of disordered region of Engrailed-2 and their contribution to DNA binding

5.1 Paramagnetic Relaxation Enhancement (PRE)

The highly flexible nature of IDPs makes it necessary to consider their "structure" as a broad ensemble of interconverting species with diverse conformations. Most IDPs possess some degree of residual structure, which we can describe as restrictions of their conformational space. The local residual structure of IDPs is defined as any conformational preference that deviates from random coil either involving certain residues to form hydrophobic clusters or secondary structure elements. Other restrictions of the conformational space involve long-range contacts (long-range is defined from the protein sequence point of view), which can be transient contacts mediated by electrostatics or some tertiary structure-like elements.[86,87] The ^{1}H-^{1}H cross relaxation (NOEs) provides a straightforward structural information in the form of interproton distances, typically up to 5 Å and is routinely used to investigate the structural features of folded proteins. The NOEs are much less informative to probe transient or persistent long-range contacts in IDPs due to very short-lived contacts between distal protons, the limitation of distance to probe contacts up to 5 Å and unfavorable timescales of local motions.[88]

The paramagnetic relaxation enhancement (PRE) is proportional to the average $\langle r^{-6} \rangle$ of the distance between a proton and the electronic paramagnetic center. The large magnetic moment of the unpaired electron renders the dipolar interaction with nuclear spins strong. The detection of long-range contacts within ~ 25 Å from the spin label is thus possible.[26] The ease to implement PRE experiments (in theory, see below) makes this technique attractive in probing the long-range contacts in IDPs with minimum perturbation to the system.

The protocol used to measure PREs in IDPs requires a single cysteine to be introduced (or kept, if the wild-type protein features cysteines) into the protein sequence at a desired position through site directed mutagenesis (all wild-type cysteines are usually mutated to serine residues).[89] In order to minimize perturbations to local residual structure, mobile regions, with "low compactness" are preferable mutagenesis sites.[17] The electron spin label, typically a nitroxide radical (like proxyl, MTSL, *etc.*) contains a group, which reacts with the cysteine side-chain thiol function under specific conditions. The high gyromagnetic ratio of the unpaired electron of the spin label makes dipolar relaxation with nuclei very efficient. The longitudinal and transverse relaxation of nuclei that are within ca. 25 Å of the spin label or transiently in closer contact are strongly affected by the presence of the spin label. The enhanced relaxation of the nuclei results in the line broadening and loss of intensity in corresponding peaks. For each proton, the ratio of intensities measured in the paramagnetic state, (when spin label is active) and the diamagnetic state (when the spin label is inactive), provides valuable information about these long-range contacts.[90–92]

Figure 36: Site-specific spin labeling of proteins with paramagnetic tags allows for the detection of long-range contacts within ~ 25 Å from the spin label. PRE resulting from magnetic dipolar interactions between a spin nucleus and the unpaired electrons of the paramagnetic center provides important information to probe long-range contacts between different regions and segments of the protein which are sequentially far way but have spatial proximity.

5.2 Design of cysteine mutants of Engrailed-2 for PRE studies

In order to investigate the conformational space of the disordered region and to map out its long-range contacts with the homeodomain of Engrailed-2, a series of cysteine mutants has been designed to introduce the paramagnetic probe 3-maleimido-2,2,5,5-tetramethyl-1-pyrrolidinyloxyl (3-maleimido-proxyl) in several sites (one at a time) in the disordered N-terminal region and at the surface of the homeodomain.

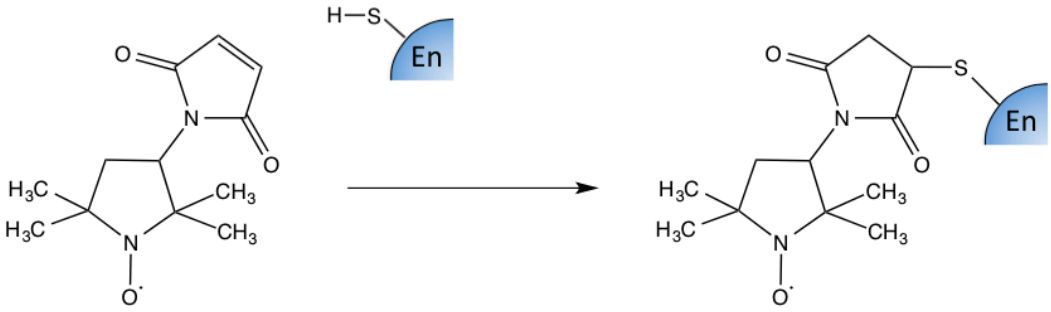


Figure 37: The chemical reaction between 3-maleimido-2,2,5,5-tetramethyl-1-pyrrolidinyloxyl (free radical) and sulfhydryl group of cysteine in Engrailed-2. The reaction leads to covalent attachment of the free radical to protein to induce paramagnetic effect.

Samples of uniformly ^{15}N labeled Engrailed-2 were incubated with 8 mM of dithiothreitol (DTT) for 30 minutes to ensure that all cysteine residues are initially in their reduced form (1 mM DTT was using in all buffers during the purification). DTT was removed by HPLC and samples were lyophilized. The lyophilized protein was then dissolved in the reaction buffer containing 50 mM sodium phosphate (pH = 6.5), 150 mM NaCl, 0.1 mM EDTA and 0.02% NaN_3. The 3-maleimido-proxyl (free radical) was dissolved in methanol and added with 10-fold excess to the protein but not exceeding more than 3% of the total volume. The reaction was allowed to proceed overnight at room temperature. Samples were then diluted to 4 times and subjected to HPLC to remove the excess of the tag. The fractions from HPLC were lyophilized and then dissolved in 40 mM succinate buffer at pH = 6.0, with a house-made antiproteases cocktail.

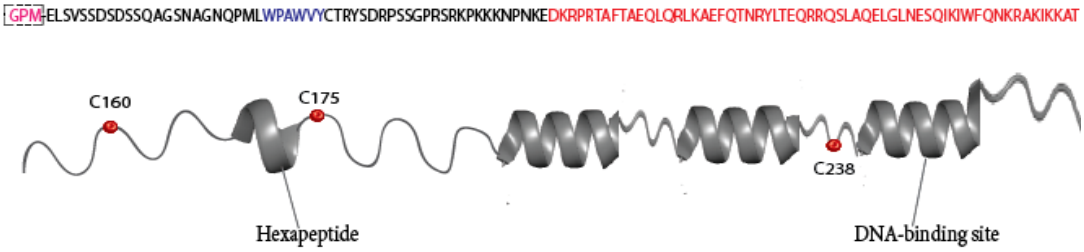


Figure 38: Sequential positions of paramagnetic tag (red) in hdn to probe long-range contacts. The presence of paramagnetic tag just at the C-terminus of the hexapeptide (C175) aims to reveal its role in establishing the long -range contacts with folded homeodomain. The placement of tag in the folded region (C238) and terminal part (C160) of the disordered region aims to understand the contribution of these parts to attain the spatially proximity with other regions of protein especially the hexapeptide.

5.3 Results

Paramagnetic relaxation enhancement effects for all constructs (see below) were measured by comparing the peak intensity of a ^{1}H-^{15}N HSQC spectrum of the active spin-label (paramagnetic state) with a reference ^{1}H-^{15}N HSQC spectrum of the diamagnetic state. The reference/diamagnetic state was obtained by reduction with ascorbic acid to quench the nitroxide in the sample used to record the ^{1}H-^{15}N HSQC spectrum of the paramagnetic state. The plot of residual intensity (I_{Para}/I_{Dia}) as a function of the sequence of Engrailed-2 is shown in Figure 39. Multiple long-range contacts between the homeodomain and the disordered N-terminal region can be identified.

5.3.1 hdn_WT (Figure 39a)

As expected, the presence of a paramagnetic label at C175 (native cysteine), just at the C-terminus of the hexapeptide, strongly attenuates the peak intensities at and around the hexapeptide. However we can also observe a strong attenuation of peaks at the N-terminus of the homeodomain and, to a lesser extent, in the two loops (α1- α2 and α2- α3) as well as the third helix of the homeodomain. These PREs results indicate the presence of some long-range contacts between the hexapeptide, where the paramagnetic tag is attached, and the globular DNA binding domain.

5.3.2 hdn_C238 (Figure 39b)

In order to confirm the proximity of the homeodomain and the hexapeptide, we decided to evaluate if a paramagnetic tag attached to the homeodomain would result in a significant signal attenuation in the disordered region and, in particular, the hexapeptide. We designed a double mutant C175S & G238C, which was tagged with proxyl. G238 is solvent exposed and located in the α2-α3 loop of the homeodomain. Sequence alignment analysis of homeodomains showed that this position is naturally substituted by a cysteine residue in several homeodomains. Thus, this proxyl tag was not expected to perturb significantly the structure of the homeodomain.

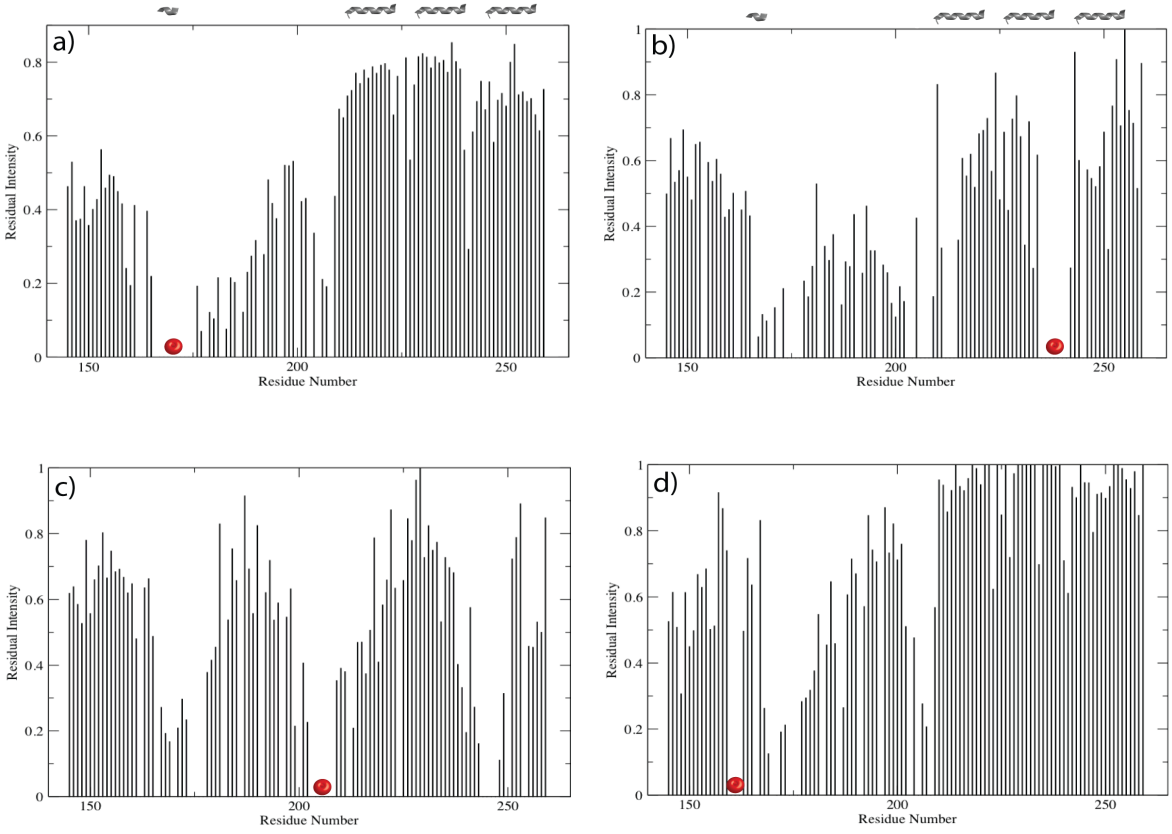


Figure 39: Plot of the PREs defined as the ratio of intensities I_{Para}/I_{Dia} as a function of the primary structure of Engrailed-2. The proxyl label was grafted at position (a) 175 on the wild-type protein; (b) the 238 position on the C175S/G238C double mutant; (c) position 205 on the C175S/T205C double mutant; and (d) position 160 on the C175S/S160C double mutant. The ratio of signal intensities (I_{Para}/I_{Dia}) reflects paramagnetic enhancement of nuclear spin relaxation caused by interaction with unpaired electron. A value close to 1 indicates no paramagnetic effect while missing lines indicate residues for which peaks could not be assigned in paramagnetic state (in most cases because the signal disappears due to very fast relaxation).

This was confirmed by limited chemical shift perturbations observed on the tagged protein (in the diamagnetic state). Signal attenuations in the disordered region and, in particular, in the hexapeptide confirmed the presence of long-range contacts between the homeodomain and the disordered N-terminal region of the protein. Additionally strong attenuation of peaks in particular also occurred in the N-terminus and in the loops of the homeodomain, as observed in the tagged wild type (C175) protein.

5.3.3 hdn_C205 (Figure 39c)

The above-mentioned results clearly indicate that the residues at the N-terminus of the homeodomain are very strongly affected by the paramagnetic effect in both cases. The proximity to the α2-α3 loop and third helix of the homeodomain is expected from the solution structure of the homeodomain (Figure 17a).[31] In order to confirm the role of this region in making the long-range contacts with both the IDR and the homeodomain of the protein, a double mutant with C175S & T205C was tagged similarly with proxyl. Strong signal attenuation was observed at the hexapeptide site and in the third helix of the homeodomain. This confirms the existence of long-range contacts between the N-terminus of the homeodomain and the hexapeptide in the disordered region of the protein.

5.3.4 hdn_C160 (Figure 39d)

In order to investigate if long-range contacts exist between other parts of the IDR and the homeodomain, a double mutant (C175S & S160C) was designed so that the proxyl radical would be at the distal part of the N-terminal tail. The signal attenuation very close to the tag was far less as compared to all other mutants. This possibly suggests the incomplete reaction of the paramagnetic tag with cysteine. Surprisingly, strong signal attenuation was measured in the hexapeptide and at the N-terminus of the homeodomain, in a pattern very similar to results obtained on the wild type (C175) protein.

Although most of these results are consistent, some PRE effects remain in some regions with similar intensity even though the proxyl tag can be grafted at different

positions of the protein. This observation may suggest the presence of free radical molecules that may transiently bind on the surface of the homeodomain or the hexapeptide. It has been observed that maleimide derivatives (Proxyl) can also attach covalently to residues other than cysteine, in particular amino group of lysines.[93] We decided to run a series of control experiments to explore the sources of potential artifacts in our protocol.

5.4 Mass spectrometry analysis

The Matrix-assisted laser desorption/ionization combined with time of flight measurements (MALDI-TOF) technique was used to analyze the yield of the reaction between proxyl and the cysteine residue. We also used this to identify the amount and type of species present in the samples used for the above-mentioned PRE experiments. For this analysis, samples were mixed with an HCCA matrix and deposited onto a metal plate. The mass of the samples was measured by a linear time of flight (TOF) mass spectrometer in the linear mode.

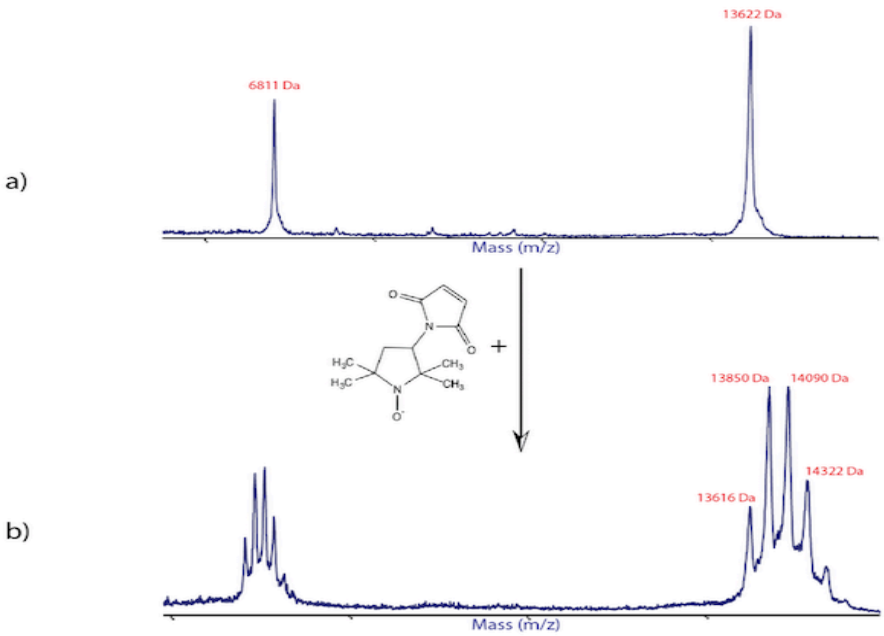


Figure 40: Mass spectrometry analysis before and after the reaction with proxyl of wild type (C175) protein. (a) The protein before the reaction is a single molecule with the expected molecular weight and does not contain any significant impurity or sign of degradation. (b) After the reaction with 3-maleimido-proxyl, multiple peaks clearly suggest that a series of sites on the protein are tagged with proxyl.

The observed weight of the untagged wild type and other mutants was in the range of 13620 - 13640 Da which was in very good agreement with the theoretical estimation (13637 Da) for the uniformly ^{15}N labeled protein (Figure 40). The reaction of the protein with proxyl is expected to increase the molecular weight of the protein by 237.28 Da. MALDI-TOF mass spectra clearly indicate the presence of a series of species of the protein with different numbers of proxyl tags (Figure 40b). The mass increment between two species corresponds very well with to the incremental mass of a single proxyl tag. Four different species that correspond to zero, one, two and three paramagnetic tags grafted to the protein show that the labeled species are not uniform and several secondary reaction sites exist in the protein. This result was surprising in so far as a low pH was used (pH 6.5) in the assay to prevent such side reaction with the nucleophilic amine group of lysines. Due to the high density of positive charges on the homeodomain, it may be possible that some lysine residues have lowered pKa and are more reactive.

5.5 Negative Control (S_175)

In order to further confirm that proxyl reacts or attaches to the residues other than cysteine, we performed the same reaction with the C175S mutant, which has no cysteine residue.

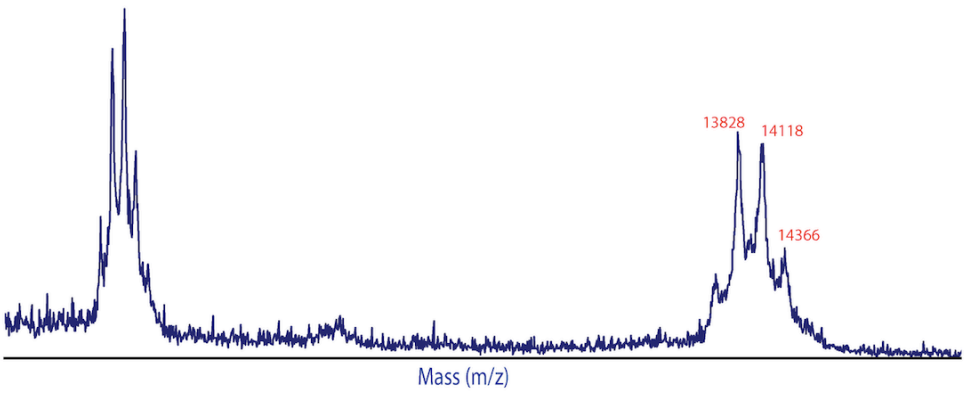


Figure 41: Mass spec analysis after the reaction with proxyl of negative control (S175) protein. After the reaction with 3-maleimido-proxyl, multiple peaks clearly suggest that a series of sites on the protein are tagged with proxyl.

MALDI-TOF mass spectra of negative control (S175) clearly indicate that proxyl reacts to residues other than cystein. The observed molecular weight shows the same increment in mass similarly to other mutants as discussed above.

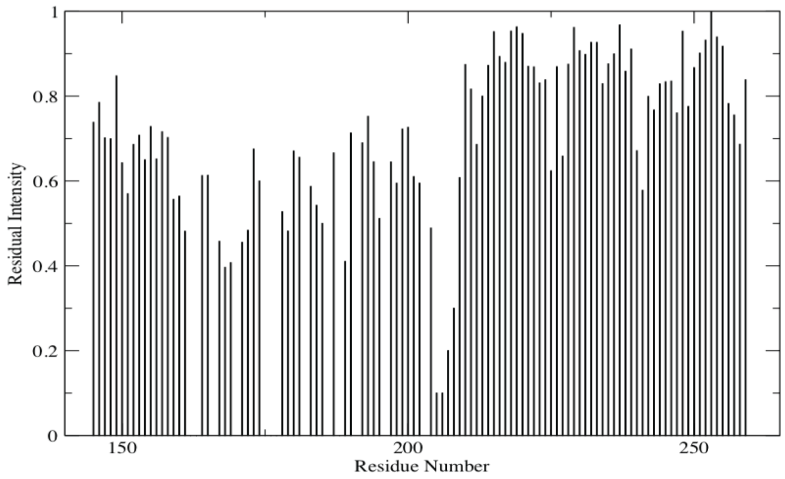


Figure 42: The plot of residual intensity (I_{Para}/I_{Dia}) as a function of the sequence of the C175S mutant of Engrailed-2, which has no cysteine residue. The observation of significant PRE effects confirms that proxyl reacts to residues other than cysteine under our conditions.

The signal attenuation with the C175S mutant confirms that proxyl reacts with residues other than cysteine. More interestingly, the behavior of signal attenuation is very similar to PREs measured with the wild type and other mutants of the protein in some regions (in particular the N-terminal part and the third helix of the homeodomain). Yet, the PRE profile of the C175S mutant is clearly different from the profiles measured with the proxyl tag grafted to a cysteine residue at various positions. Long-range contacts between disordered region and homeodomain are plausible but a system with a better control of the labeling site (a different chemistry) has to be used to allow unambiguous results and their quantitative analysis of potential long-range contacts between the disordered region and homeodomain.

5.6 MTSL (an alternative tag)

MTSL (*S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfono-thioate) is the most widely used nitroxide paramagnetic spin label to record PREs. It has been extensively used in the investigation of residual structure in

IDPs.[94,95] The reaction with the thiol group of a cysteine side chain leads to the formation of a disulfide bond. A single MTSL tag adds 186.3 Da to the mass of the protein.[96] The specificity of the chemistry and the small molecular volume (similar to natural tryptophan or phenylalanine side chains) of MTSL are major advantages compared to maleimide derivatives (Proxyl) that can also modify the amino group of lysines.

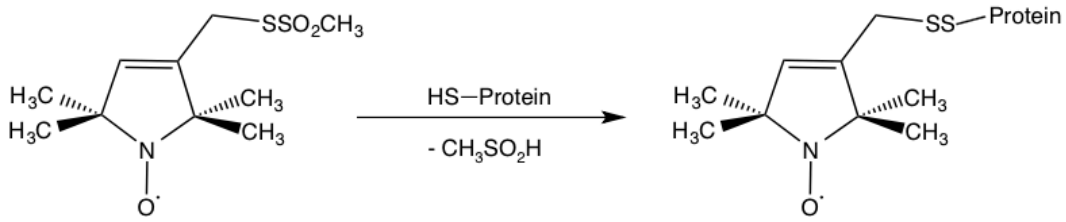


Figure 43: The chemical reaction between free radical MTSL (*S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate) and the thiol function of a cysteine side chain in Engrailed-2. The reaction leads to a covalent disulfide bond between the nitroxide radical and the protein to induce paramagnetic relaxation effects.

PRE experiments were repeated by using MTSL as the paramagnetic spin label. A protocol similar to the one used for proxyl was employed for the reaction between MTSL and a cysteine residue. The only difference was the use of pH = 7.0 instead of 6.5 as literature shows that this pH gives the maximum yield for the reaction. Samples were analyzed with MALDI-TOF technique to calculate the yield of the reaction and to investigate whether MTSL may also lead to secondary reactions in a way similar to proxyl.

The MALDI-TOF mass spectrum after reaction shows an increase of the molecular mass upon binding equal to 184 Da, which is very close to the expected 186.3 Da for a single MTSL label. MTSL reacts only to the single cysteine residue, unlike proxyl, and the yield of the reaction is close to 100%. These samples could then be used to record PRE experiments.

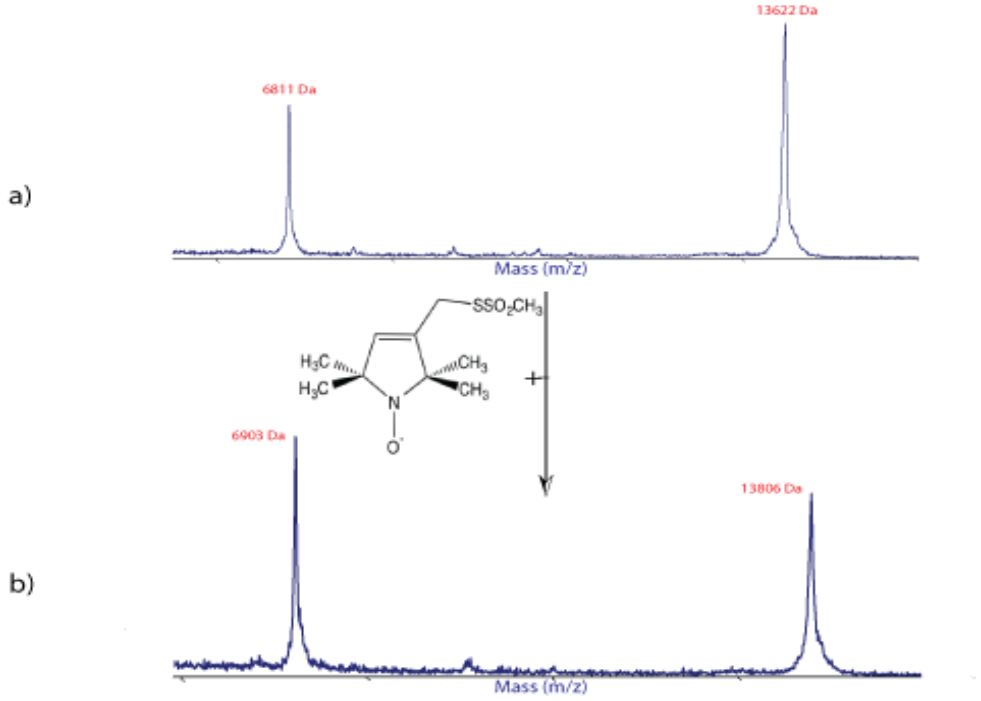


Figure 44: Mass spectrometry analysis of wild type (C175) protein sample used for PREs before and after reaction with MTSL shows that (a) the protein before the reaction is single molecular species with the expected molecular weight and does not contain any significant impurity or sign of degradation. (b) The mass spectrum of the product of the reaction with MTSL shows a single peak with a mass increment that corresponds to the addition of a single MTSL tag. MTSL is cysteine specific under these conditions and the yield of the reaction approaches 100%.

5.7 Results

Although it is clear from the MALDI-TOF analysis that MTSL specifically binds to a single cysteine, the PRE results show very similar behavior of long range contacts compared to the results with proxyl. This reassuring result shows that there are indeed some long-range contacts between the IDR and the homeodomain of Engrailed 2.

5.7.1 hdn_WT (Figure 45a)

The presence of paramagnetic label at C175 (native cysteine), just at the C-terminus of the hexapeptide, strongly attenuates the peak intensities at and around the hexapeptide, which confirms that the reaction was complete. Similarly to the

observation with the proxyl tag, strong attenuation of peaks in particular occurs at the N-terminus (200-208) of the homeodomain, which is a flexible region. The other loop regions (α1-α2 222-227, α2-α3 238-242) between the helices of the homeodomain also show some degree of proximity to the MTSL tag. In addition to these flexible linkers, a few residues in the first and third helices (215-218, 247-253) show significant intensity attenuation. This suggests that some accessible states in the conformational space of Engrailed have the IDR in contact with an important proportion of the surface of the homeodomain, in particular, and most interestingly, part of the DNA binding surface (discussed below).

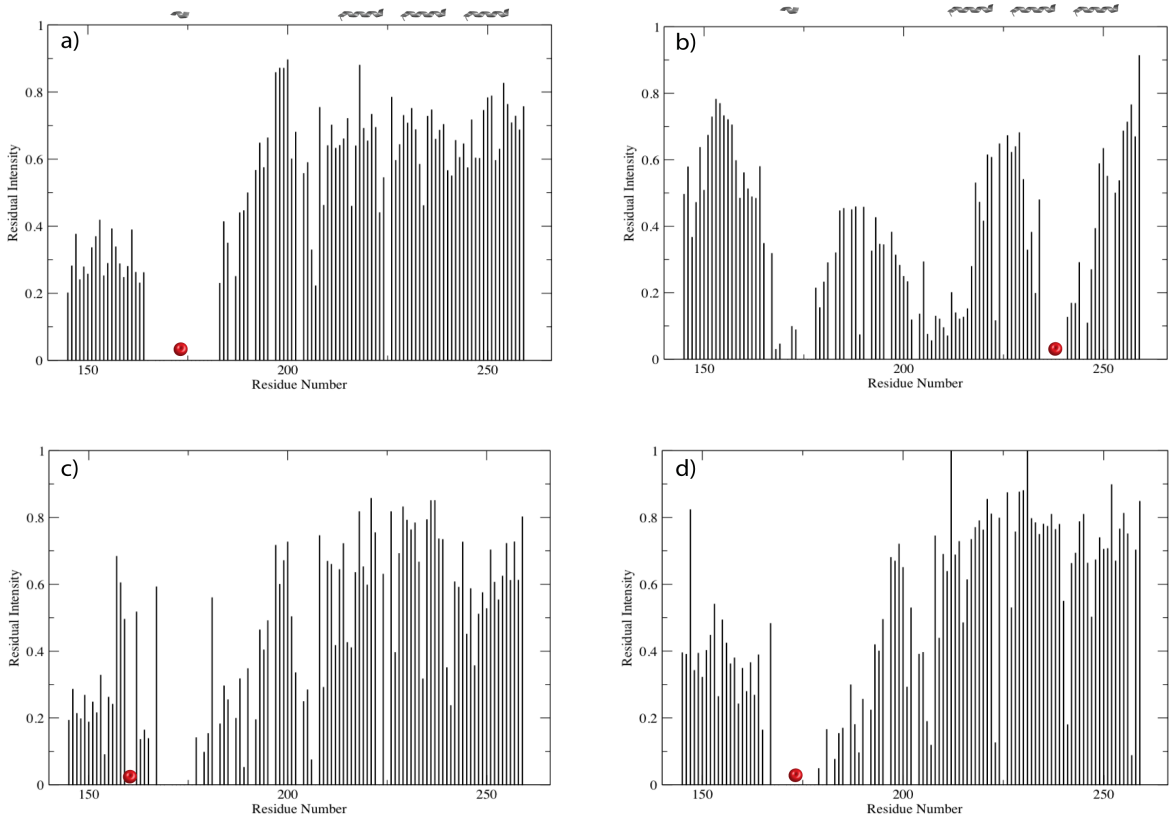


Figure 45: Paramagnetic relaxation effects (I_{Para}/I_{Dia}) as a function of the sequence of Engrailed-2. The MTSL label was grafted at position (a) 175 on the wild-type protein; (b) the 238 position on the C175S/G238C double mutant; (c) position 160 on the C175S/S160C double mutant; and (d) position 175 of the double mutant W169A/W172A.

5.7.2 hdn_C238 (Figure 45b)

In order to confirm the proximity of the homeodomain and the hexapeptide, we decided to evaluate if a paramagnetic tag attached to the homeodomain would result in a significant signal attenuation in the disordered region and, in particular, the hexapeptide. We used the double mutant C175S & G238C, which was tagged with MTSL. This position clearly attenuates the signals around the paramagnetic tag and at the N-terminus of the homeodomain, which is expected from the 3D structure of the homeodomain. Importantly, some parts of the IDR, in particular, around the hexapeptide (residues 165-180) show significant PREs, which confirms the presence of long-range contacts between the homeodomain and the disordered N-terminal region of the protein. Interestingly, significant PREs are detected on the disordered region between the hexapeptide and the homeodomain and not on the disordered region at the N-terminus of the hexapeptide. This suggests that the part of the IDR of Engrailed that covers partially the surface of the homeodomain ranges from the N-terminus of the homeodomain to the hexapeptide but not beyond.

5.7.3 hdn_C160 (Figure 45c)

In order to investigate if long-range contacts exist between other parts of the IDR and the homeodomain, the double mutant C175S/S160C was designed, so that the nitroxide radical would be at the distal part of the N-terminal tail. The very close region to position C160 was least affected compared to other positions. The tag at this position also affects very strongly the hexapeptide and other flexible linker regions similar to other tag positions. But the marked difference was that almost all residues in the third helix (243-253) showed small but significant, which hints about the possibility of N-terminal tail to occupy spatial proximity to third alpha helix.

5.7.4 hdn_2W/A_C175 (Figure 45d)

In order to identify the role of hydrophobic residues in the hexapeptide and their contribution to long-range contacts, two tryptophans of the hexapeptide were mutated to alanines (2W/A) to disrupt the hydrophobic cluster. The W169A/W172A

double mutant was labeled with MTSL on the native Cys175 residue. If hydrophobic contacts were mediating the interaction of the hexapeptide and the homeodomain dramatic changes in the PRE profiles should be expected. This was not the case. Results on this double mutant are similar to those obtained on the wild type protein. This suggests that long-range contacts of the hexapeptide and the homeodomain are independent of the hydrophobicity of the cluster.

5.8 Mapping of interactions on homeodomain

All paramagnetic relaxation enhancements indicate that the disordered region occupies a volume that covers regions of the homeodomain in at least part of the conformational ensemble of Engrailed. The residues of the homeodomain that show a significant PRE with a MTSL tag attached to the IDR were mapped on the solution structure of the homeodomain (pdb code 3ZOB)[31] (see Figure 46a).

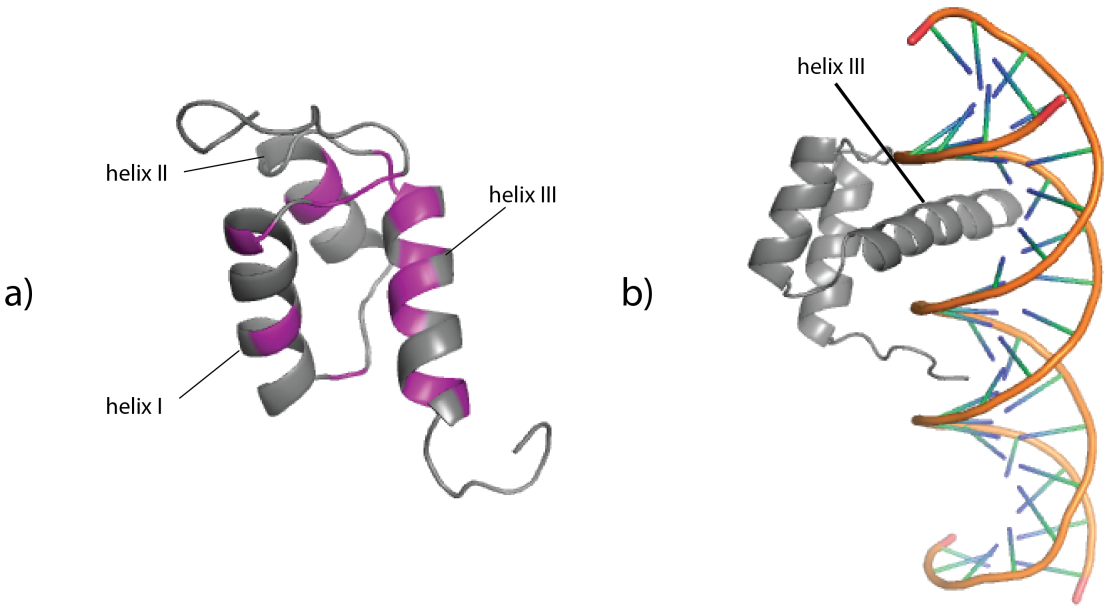


Figure 46: (a) The residues in homeodomain highly affected (50% and more) by the long-range contacts of disordered region are highlighted on the NMR structure of the homeodomain of Engrailed-2 (pdb code 3ZOB).[31] The third helix is at least partially covered in some conformers with significant populations. (b) The crystal structure of the homeodomain of Engrailed bound to DNA solved by Kissinger *et al.*[32] shows that the third helix fits into the major grove of the DNA (pdb code 1HDD).

The X-ray structure of the complex of the homeodomain of Drosophila Engrailed with DNA was solved by Kissinger *et al.*[32] and is shown in Figure 46(b). The third helix binds to the major grove while the residues at the N-terminus of the homeodomain fit into the minor grove of this duplex DNA. Helices 1 and 2 face each other in an antiparallel arrangement and are not in direct contact with DNA.[32] Clearly, the region of the homeodomain that is in contact with the disordered region overlaps significantly the surface of the homeodomain that is involved in the interaction with DNA. Interestingly, very few contacts were identified between the IDR and the surface common to helices α1 and α2, which faces away from the DNA.

The PREs measured on the third helix of the homeodomain are slightly stronger when the nitroxide label is grafted to residues 160 of the IDR than when it is grafted at position 175. This seems to indicate that the IDR at the N-terminus of the hexapeptide is in closer contact to this part of the surface of the homeodomain. On the other hand, the nitroxide at position 160 does not lead to paramagnetic relaxation enhancements in loop α2-α3. This is in good agreement with the fact that the nitroxide label at position 238 leads to limited relaxation enhancements in the IDR around position 160. Thus, we can conclude that the IDR at the N-terminus of our construct is specifically in contact with the surface of the homeodomain around helix 3.

The paramagnetic tag at position 238 leads to significant PREs throughout the IDR between the hexapeptide and the homeodomain and at the hexapeptide. This seems to indicate that this part of the disordered region of Engrailed crowds the surface of the homeodomain near loop α2-α3. Surprisingly, no significant PRE could be detected on residue Gly238 when the nitroxide label is grafted to position Cys175. This could be due to the fact that the IDR is at a distance of the surface of the homeodomain and is more likely to be close to the nitroxide label than to the homeodomain itself.

The contribution of disordered regions of Antennapedia homeoprotein to DNA binding studied by Quian *et al.* demonstrates that C- and N- terminal disordered regions reduce the affinity of the protein to DNA.[53] In particular, the specificity of the binding of the Labial homeoprotein (Lab) to DNA requires the interaction of its hexapeptide to the partner homeoprotein Exd to form a ternary complex. Mutation of

the hexapeptide in Lab from YKWM to AAAM alleviates the need for Exd to bind DNA.[43] These studies clearly indicate that the hexapeptide of Lab has an auto-inhibitory effect in the absence of other partner proteins. This auto-inhibitory effect is possibly due to the competitive interaction of the hexapeptide to the DNA binding sites. Partner homeoproteins would then bind the hexapeptide region and uncover the DNA binding surface (helix 3 & extended N-terminal arm of homeodomain), which becomes available to bind major and minor grooves of the DNA. Similar observation has been recently documented for conformationally disordered serine-rich region (SRR) of the E26 transformation-specific (Ets-1) transcription factor that transiently interacts with its DNA-binding ETS domain resulting in the auto-inhibition of the Ets-1 transcription factor.[97]

The existence of such auto-inhibitory effects of the hexapeptide in Engrailed-2 could be correlated to the interaction between the hexapeptide and the globular DNA binding domain. However, the concomitant mutation of both tryptophan residues in the hexapeptide seems to have more of a slight shift of balance effect rather than a dramatic change in the IDR-homeodomain interactions. This could be due to the fact that (i) the interaction between the IDR and the homeodomain in Engrailed is very weak, (ii) this interaction is not mediated by the aromatic residues of the hexapeptide or (iii) that the mutation of only two tryptophan residues in Engrailed is not sufficient to alter this interaction significantly. Note that these hypotheses are not mutually exclusive.

In order to further investigate the DNA binding properties of the globular homeodomain and the role of hexapeptide in the modulation of DNA binding, fluorescence anisotropy measurements were carried out to assess the affinity of different Engrailed-2 constructs to DNA, with or without its disordered region.

5.9 Fluorescence anisotropy

Fluorescence anisotropy is a technique that can be used to measure the thermodynamic constants of binding equilibra. The principle of this method is to tag one of the interacting partners with a fluorophore. A polarized excitated light applied to fluorophores yields to emission light of different intensities according to the orientation of the fluorophore. The degree of polarization of the emitted light is related to the correlation time τ_C of the fluorophore according to the equation:

$$r = r_0/(1 + \tau/t_C) \qquad (19)$$

where r is the observed anisotropy and depends on the intensity difference of parallel and perpendicular emission components of the polarized light with expression $(\frac{I_{\parallel}-I_{\perp}}{I_{\parallel}+2I_{\perp}})$, r_0 is the limiting intrinsic anisotropy of the fluorophore at t = 0, τ is the fluorescence limetime.[98]

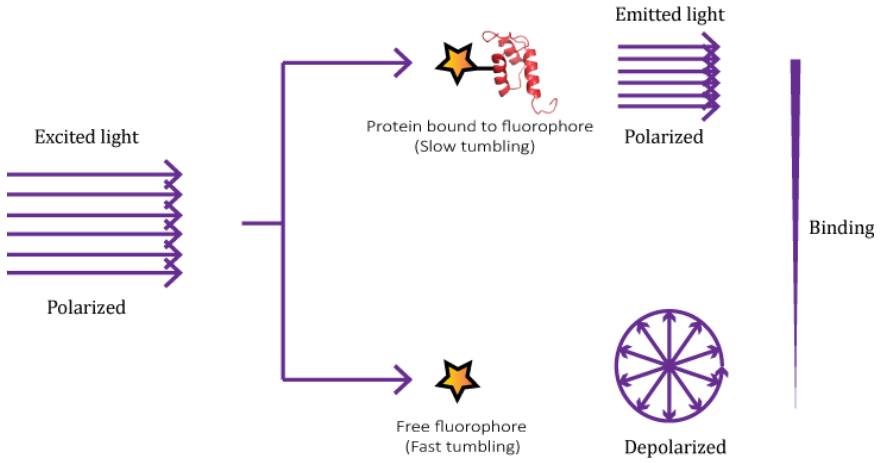


Figure 47: The principle of fluorescence anisotropy for measuring the binding interaction between two molecules. The extent of tumbling of fluorophore covalently attached to a ligand excited by polarized light is measured to determine the association constant of interacting molecules. Fluorophore bound to the ligand in the complex tumbles slowly and emits polarized light while unbound fluorophore will tend to scramble the polarized light.

If the molecule containing the fluorophore is unbound to the partner molecule then its tumbling will be faster and emitted light will be less polarized as compared to the molecule bound to partner. The difference in anisotropy between bound and unbound states enables the straightforward determination of the K_d of the interaction.

A duplex DNA of 20 bp was synthesized. The chosen sequence corresponds to a known Engrailed binding region derived from the HF2 region of MAP1B promoter.[99] Engrailed binds to the consensus TAATAT sequence.[100] The two single strands have the following sequences;

$^{5'}$*CTAGCATAATATGTCTGCCT$^{3'}$

$^{3'}$GATCGTATTATACAGACGGA$^{5'}$

An Alexa Flour® 488 fluorophore was attached to the 5'-end of the first strand (*). For hybridization, the two strands were mixed in equal amount and incubated at 90 °C for 3 minutes and left overnight for steady cooling. The predicted T_m of the duplex is 47 °C with a GC content of 40%.

A fixed nanomolar concentration of DNA was titrated with increasing concentrations of protein (nM-μM range) dissolved in 25 mM HEPES buffer (pH = 7.4) and 150 mM KCl to prevent non-specific electrostatic binding to DNA. After mixing the protein and DNA in a 96-well reader plate, the plate was placed in dark for minimum 3 hours for equilibration. Fluorescence anisotropy was measured at 25 °C, using a 485/520 nm fluorescence polarization module. The Hill equation[101] was used to determine the K_d of interaction.

$$Y = r_{max} \times \frac{X^h}{(K_d^h + X^h)} \tag{20}$$

where r_{max} is the maximum anisotropy upon saturation, K_d is the ligand concentration required to achieve half-maximum binding and h is the hill slope.

5.10 Different constructs bind with different binding affinity

The dissociation constants (K_d) of different Engrailed-2 constructs were measured (Figure 22): wild type hdn (including the homeodomain and disordered region), homeodomain with mutated hexapeptide in the disordered region (hdn_2W/A), homeodomain without hexapeptide (nls) and homeodomain only (hd).

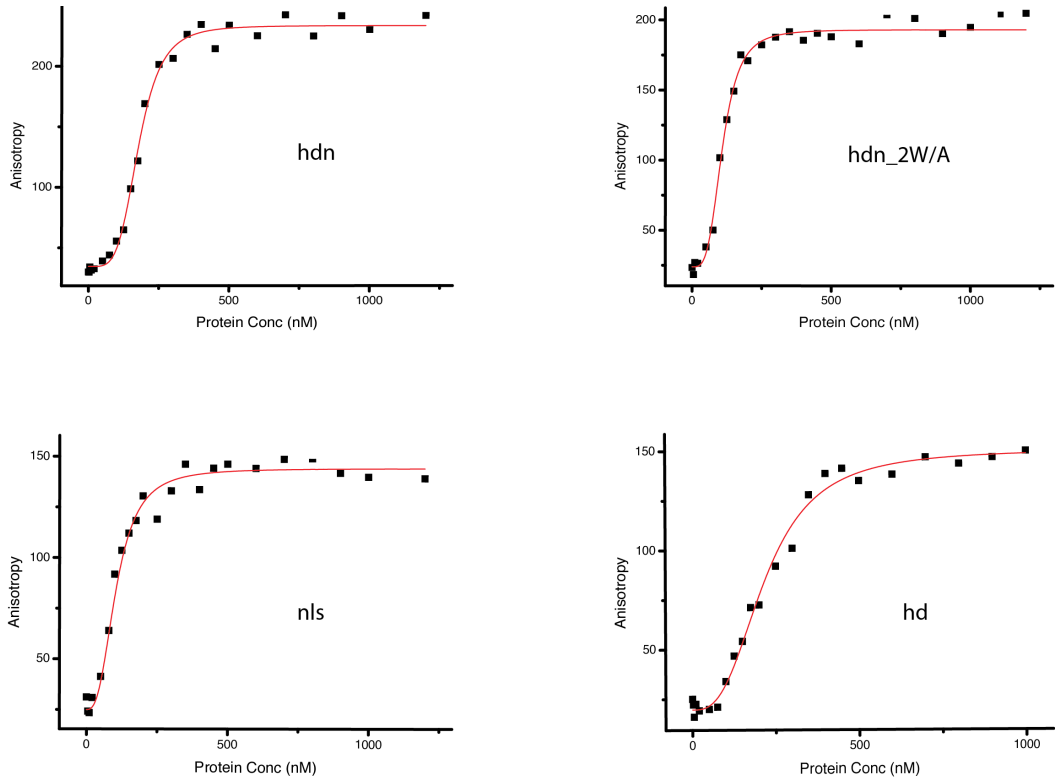


Figure 48: Saturation binding curves of a duplex DNA labeled with Alexa fluorophore binding to different Engrailed-2 constructs (hdn, hdn_2W/A, nls, hd). Experimental points were fitted with a Hill equation (red lines).

The dissociation constants found for each Engrailed-2 constructs are summarized in Table 2 below. These dissociation constants clearly indicate that all these constructs bind with a different affinity to DNA. This observation suggests the possible contribution of disordered region in modulating the DNA binding affinity. The higher binding affinity of hdn_2W/A as compared to hdn suggests that the hexapeptide has an auto-inhibitory effect in the absence of partner protein. However our results do not show a marked difference in the binding affinity of hdn and hdn_2W/A to DNA. For the homeodomain, the results were contrary to our hypothesis that

homeodomain alone will bind more strongly to DNA as transient auto-inhibitory effect from hexapeptide does not exist any more. However the lowest K_d was observed for the nls, which is an elongated homeodomain construct lacking the hexapeptide region. This suggests that additional positive contributions from the N-terminal tail of homeodomain may be involved in DNA binding. Examination of Engrailed-2 sequences shows the presence of a conserved segment enriched with basic residues (KPKKKNPNK) upstream of the homeodomain (residues 190-198). This basic segment is a likely candidate to synergize homeodomain DNA binding.

Construct	K_d (nM)
hdn	177
hdn_2W/A	108
nls	104
hd	217

Table 2: Summary of the dissociation constants (K_d) of various Enfrailed-2 constructs for the fluorophore tagged DNA.

Overall, these results indicate that all dissociation constants were in the 10^{-7} M range, which is higher than the expected K_d according to the previous work of Kissinger *et al.*, describing affinities of homeodomain for DNA in the 1-2 nanomolar range.[32]

5.11 Competition assay

One possible reason for higher K_d could be that the fluorophore tag at 5'-end of the DNA, which is not very far away from the interaction site may possibly interfere with binding by partially obstructing the interaction site for the protein. To address this problem, we chose to use a competitive-binding assay to determine the K_d of the interaction between protein and untagged DNA. Competitive-binding assay is a method in which an unlabeled biologically specific binding agent competes with labeled compound. In this case the binding between protein and labeled DNA will

result in a high anisotropy value. The addition of unlabeled DNA with stronger affinity will compete for the binding site and dissociate labeled DNA from the protein thus leading to a decrease of the anisotropy value.

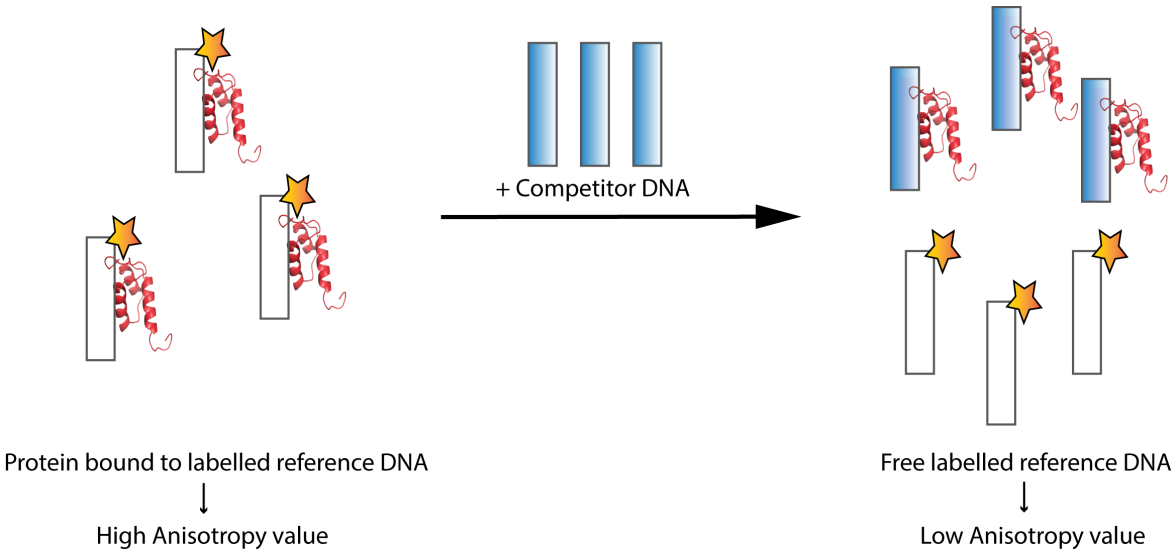


Figure 49: The schematic representation of the competitive binding assay. Engrailed-2 binds to labeled reference DNA resulting in a high anisotropy value. The unlabeled DNA with higher affinity to Engrailed-2 will compete with labeled reference DNA for binding to Engrailed-2. The substitution of labeled reference DNA will result in lower anisotropy value.

For competitive-binding experiments, 20 nM labeled DNA was mixed with the different Engrailed-2 constructs. After 3 hours of equilibration, increasing concentrations of the unlabeled DNA (10, 20 and 30 nM) were added to compete with labeled DNA for binding to the proteins. Fluorescence measurements were carried out after 16 to 24 h of incubation at room temperature.

5.12 Fluorophores potentially interfere with binding

The gradual shifts of the titration curves show that unlabeled DNA is able to displace labeled DNA from preformed Engrailed-2/DNA complexes. Using a competitive binding equilibrium, K_d of unlabeled DNA could be derived from the measured apparent K_d of labeled DNA using the following equation[102]

$$K_{d\ app}^{DNA*} = K_d^{DNA*}(1 + \frac{[DNA]}{K_d^{DNA}}) \quad (21)$$

For all investigated constructs, lower values of K_d were found with unlabeled DNA in the competitive-binding assay. This indicates that the unlabeled DNA has a higher affinity compared to reference labeled DNA. Therefore the covalent attachment of Alexa-488 fluorophore to DNA hinders the binding site of the DNA for protein.

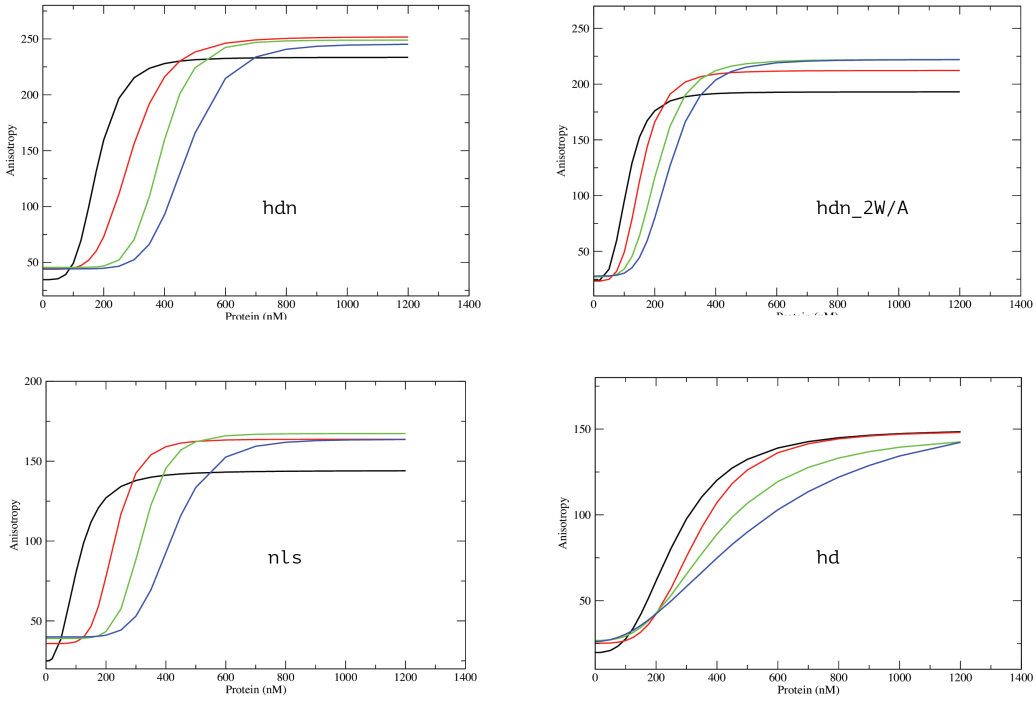


Figure 50: The competition of the Alexa 488-labelled DNA with unlabeled DNA results in a right shift of the saturation curves. The parallel shift of the binding curve (fitted values) upon adding the final 10 (red), 20 (green) and 30 (blue) nM concentration of unlabeled DNA for various constructs of Enfrailed-2 enables to determine the K_d for unlabeled DNA.

The Engrailed 2 construct encompassing the homeodomain and the disordered region (hdn) binds to DNA an almost order of magnitude higher with a K_d of = 19 nM. A similar 10 fold stronger binding was found for the homeodomain without hexapeptide (nls, K_d 11 nM) and for homeodomain only (hd K_d 27.5 nM). But in the case of the homeodomain with mutated hexapeptide (hdn_2W/A K_d= 23 nM) the affinity was only 5 fold higher.

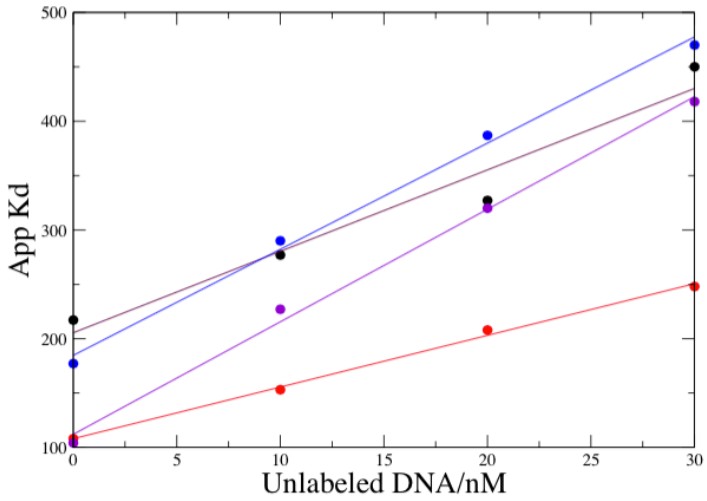


Figure 51: Plots of the apparent Kd of labeled DNA for hdn (blue), hd_2W/A (red), nls (indigo) and hd (black) constructs as a function of the concentration of unlabeled DNA.

The value of K_d for homeodomain with mutations in the hexapeptide (hdn_2W/A) compared to wild type hdn was also found to be higher in the competitive-binding assay. This observation contradicts the affinity values obtained when direct binding was used to measure the K_d as summarized in Table 3.

The nls construct has a higher affinity for DNA in comparison with hd only, confirming that the N-tail region makes a positive contribution to binding.

Construct	**Direct Binding** K_d(nM)	**Competitive Binding** K_d(nM)
hdn_WT	177	18.91
hdn_2W/A	108	22.73
nls	104	10.8
hd	217	27.5

Table 3: Comparison of dissociation constants (K_d) for various constructs of Engrailed-2 measured by direct and competitive methods.

Overall, all Engrailed-2 constructs bind to DNA with a strong affinity and no marked differences in the values of K_d were found. But still different K_d values clearly suggest that the role of disordered region especially the hexapeptide in DNA binding cannot be ignored. Our results indicate that the disordered region contributes both positively and negatively to DNA binding.

On the basis of NMR findings that have been discussed in the previous sections, it is evident that hdn possesses well folded homeodomain comprising three alpha helices connected by short loop regions and hexapeptide in disordered region that has some structural propensity as shown in Figure 52.

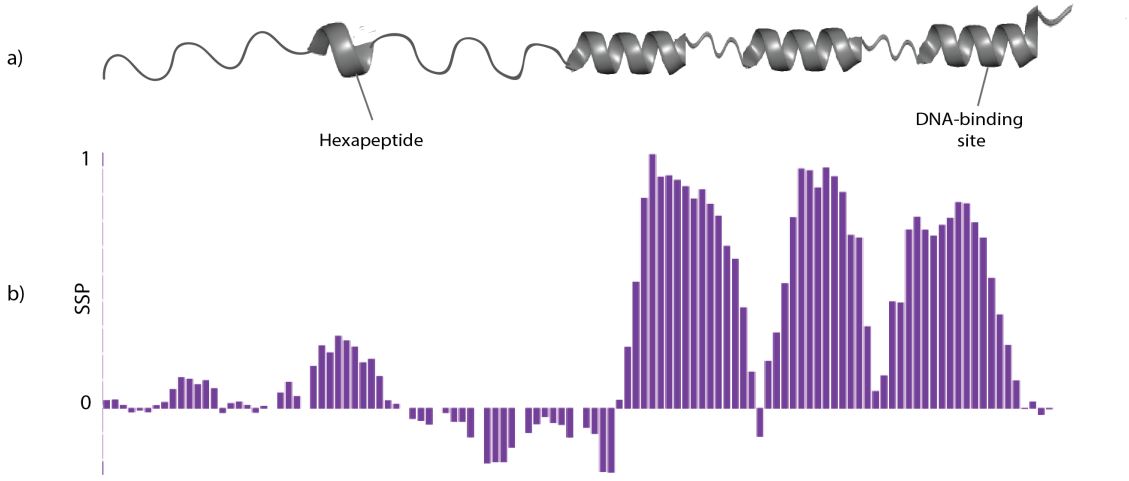


Figure 52: (a) Schematic image of the construct (hdn) under study showing the well-folded homeodomain comprising three alpha helices connected by short loop regions and hexapeptide in disordered region that has some secondary structure propensity. (b) SSP score of hdn construct clearly indicating the well-folded homeodomain comprising three alpha helices while hexapeptide showing SSP score higher than other disordered region.

The structural propensity existing at hexapeptide in disordered region somehow makes question mark that either these structural contents are locally populated or they come through long-range contacts with the folded homeodomain. The PRE results clearly indicate that long-range contacts exist between hexapeptide and homeodomain. So to further investigate the structural existence of the hexapeptide and its interaction with homeodomain, we used EPR (Electron Paramagnetic Resonance) to investigate the structural contents of the hexapeptide.

5.13 Electron Paramagnetic Resonance (EPR) Spectroscopy

EPR is a spectroscopic technique, which detects species that have paramagnetic centers/unpaired electrons. It is also often called ESR (Electron Spin Resonance). EPR is a magnetic resonance technique, the electronic equivalent of NMR. Unpaired electrons can be found in many chemical and biological systems. EPR is a method of choice to detect and identify free radicals or some ions of transition metals such as iron, copper, and nickel. The conclusive involvement of free radicals in pathological diseases intensifies its importance in free radical research, especially in radiation biology-based experimentation.[103]

The introduction of paramagnetic center in proteins through site-directed spin-labeling (SDSL) at selected site makes it a powerful technique for detecting structural and dynamical changes in proteins. A cysteine is introduced into the protein sequence at a desired position through site directed mutagenesis.[89] A low compactness value is usually preferable criteria for the labeling site to avoid any significant perturbation to the system.[17] A tag is introduced in a chemical reaction under specific conditions between the cysteine and a thiol-reactive nitroxide (like proxyl, MTSL, *etc.*). SDSL EPR spectroscopy has emerged as an important technique for mapping secondary structure elements in a wide range of proteins.

EPR is a sensitive method to map out the conformational changes arising due to protein-protein or protein-ligand interactions.[104] The monitoring of mobility and relaxation properties of the spin label enables this technique to characterize interaction and overall structural properties of the proteins.[105] The introduction of two spin labels in the same protein enables this technique to measure distances between two spin labels between 1.5 to 8 nm.[106] Such long-distance measurements provide important information and are extremely complementary to the short distance measurements possible by NMR.

When a nitroxide radical with an unpaired electron is placed in a magnetic field, the interaction of the unpaired electron with the magnetic field leads to a Zeeman effect (like nuclei with a non-zero spin): the electronic energy levels split in two different energy levels creating two spin states, Ms = ± ½. The energy difference is given by,

$$\Delta E = g\beta B$$

where g is a dimensionless constant ("g-factor"), β is the unit magnetic moment of the spinning electron called the Bohr magneton and B is the external magnetic field.

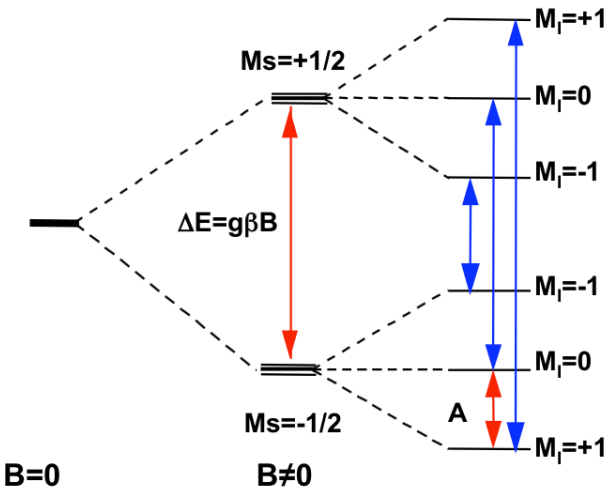


Figure 53: Diagram of the energy levels for an electron spin and I = 1. Δ*E* is the energy splitting caused by the Zeeman interaction due to external magnetic field. *A* is the hyperfine coupling constant. Blue arrows show the three possible transitions. Figures are reproduced from Ref.93.[93]

For a nitroxide radical, the hyperfine interaction of the electron spin with the nuclear spin of ^{14}N (I=1) subdivides these energy states into three different energetic levels corresponding to M_I (M_I = -1; M_I = 0; M_I = +1). The resonance condition is achieved by applying electromagnetic radiation at a suitable frequency υ that matches the energy difference Δ*E* between two levels. The EPR spectrum (in continuous wave mode) is recorded by changing the magnetic field while keeping the electromagnetic radiation frequency υ constant.

In the isotropic regime, when the radicals are highly mobile, the spectrum is composed of three equidistant lines coming from the averaging of the hyperfine interaction (Figure 54a). In a frozen solution, the anisotropy of the hyperfine interaction is revealed giving a spectrum with well-separated $M_I = +1$ and $M_I = -1$ components as shown in (Figure 54b).

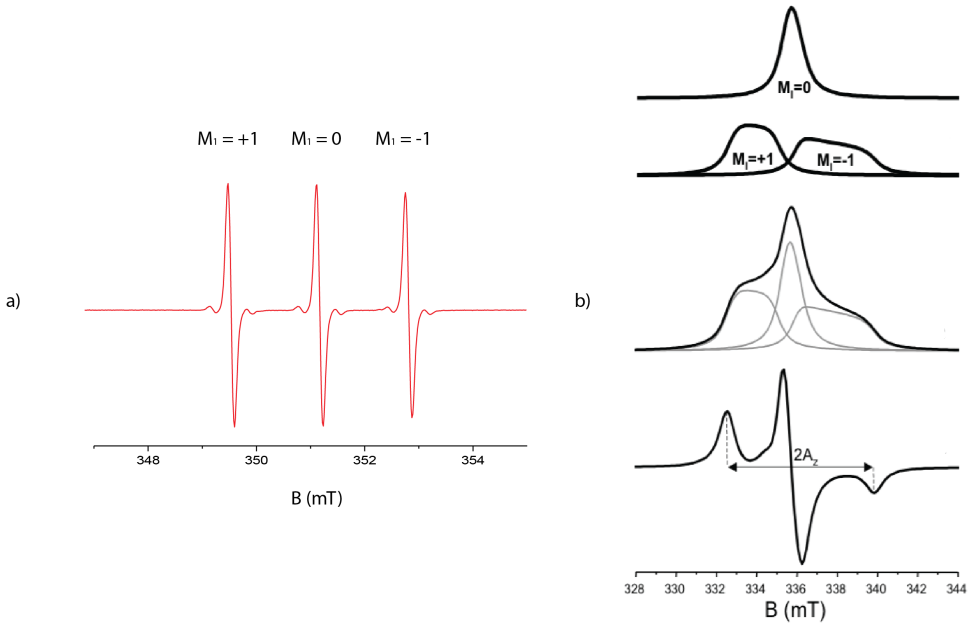


Figure 54: (a) EPR absorption lines of a nitroxide radical in the isotropic regime are composed of three equidistant lines. (b) EPR absorption lines of the same radical in frozen solution (top) and EPR spectrum as it is acquired experimentally (bottom) Figures are reproduced from Ref.93.[93]

The line shape of the spectrum characterizes the mobility of the nitroxide radical. The fast motional regime results in line broadening due to prevalence of some degree of magnetic anisotropy, while line positions remain constant (Figure 55A). Intermediate motional regime results in shape distortions of the EPR spectrum due to prevalence of high degree of magnetic anisotropy (Figure 55B). While slow motional regime results in spectra similar to frozen state (Figure 55C) as full effects of the anisotropy prevails.

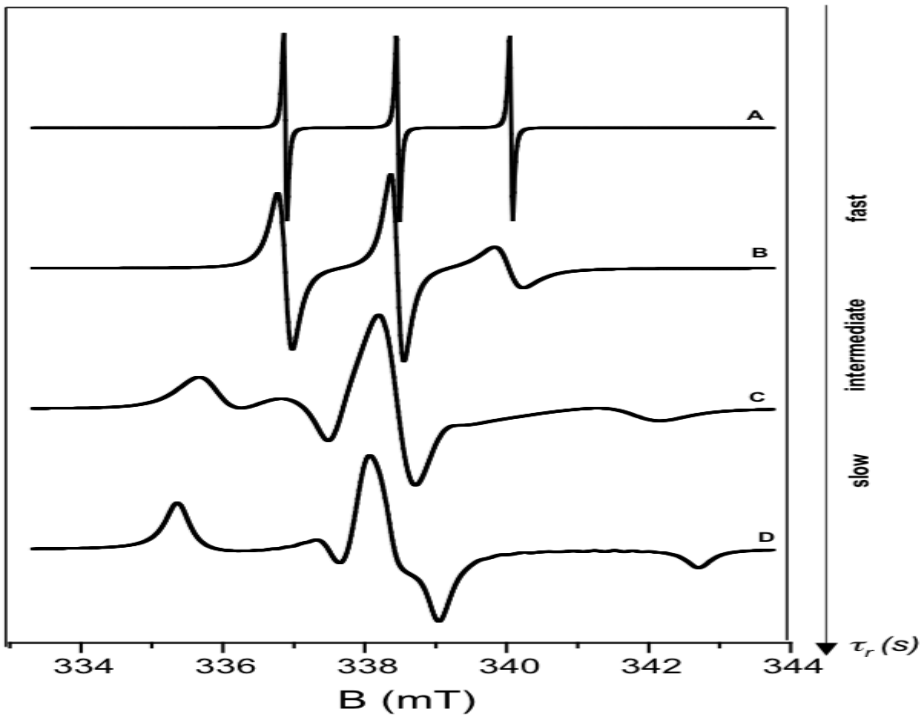


Figure 55: Set of various motional regimes and the corresponding characteristic EPR spectra for isotropic motion of a nitroxide spin label described by the isotropic rotational correlation time τ_r. The fast regime holds for τ_r < 1 ns; the intermediate regime corresponds to1 ns < τ_r < 1 μs; and the slow regime to > 1 μs. Figures are reproduced from Ref.93.[93]

The spin label attached covalently to a protein is useful for tracking the chemical environment around the spin label. The relationship between the mobility of spin label and the protein structure characterizes well the movement of the entire protein, transiently existing tertiary contacts and internal dynamics of the protein.[107]

When considering the dynamics of a nitroxide label grafted to a protein, rotational diffusion of the protein (with correlation time τ_{prot}) and the local motion of the nitroxide label (effective correlation time τ_{local}) have be taken into account. An effective correlation time τ can be defined as,

$$\frac{1}{\tau} = \frac{1}{\tau_{prot}} + \frac{1}{\tau_{local}} \qquad (22)$$

In a large protein ($\tau_{prot} \gtrsim 20$ ns), the contribution of τ_{prot} is often negligible if τ_{local} is close to 1 ns or below. In order to maximize the effects of local protein dynamics at

the labeling site, the overall tumbling can be slowed down by the addition of sucrose for instance.

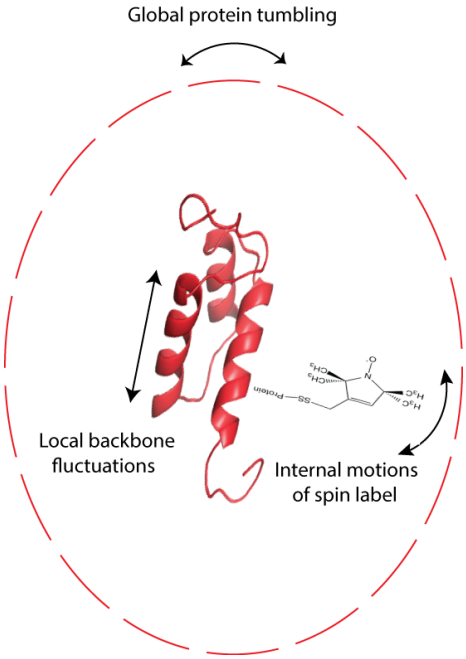


Figure 56: Schematic representation of the different contributions of mobility that influence the EPR spectral shapes of the spin label.

5.14 EPR analysis of Engrailed-2

Here, we want to address a series of questions: how different are the motions of the hexapeptide site and of the homeodomain and how are the motions of the hexapeptide altered by the presence of the homeodomain. In particular, we would like to confirm or infirm the hypothesis of a transient interaction between the hexapeptide and the homeodomain. Three constructs of Engrailed-2 were used to investigate this hypothesis: the wild type protein, hdn_C175, the double mutant hdn_C238, with a single cysteine at position 238 and a short construct limited to the disordered region (residues 146-199) without the folded homeodomain (hdn_n_C175) (Figure 22). All proteins were tagged with MTSL using a protocol described in detail below.

The lyophilized protein was dissolved in 200 µL, 40 mM sodium succinate (pH = 6). Before spin labeling, dithiothreitol (DTT) was added to protein in a molar excess of 1: 100. The mixture was incubated for 30 min to one hour at room temperature to reduce free cysteine residues. DTT was removed by PD10 desalting column (GE Healthcare) using sodium succinate 40 mM (pH = 6) as elution buffer. The fractions containing the protein were pooled. Spin label (MTSL) was immediately added to the sample at a molar excess of 10:1 using a spin label stock solution at 38 mM in acetonitrile. The reaction was carried out for 1 h in the dark in an ice bath, under gentle stirring and a continuous flow of argon to avoid oxidation. In the case of hdn_C238, three additions of MTSL were used, with one hour between each. The excess of unbound spin label was removed by gel filtration as described above, using the same buffer as previously. The fractions giving an EPR signal of labeled proteins were pooled and concentrated by ultrafiltration using a 3 kDa cutoff cellulose membrane (Amicon, Millipore). The yield obtained was ranging from 20 tot 80 %.

The EPR spectra for hdn_C175, hdn_C238 and hdn_n_C175 were recorded at room temperature on an ESP 300E Bruker spectrometer equipped with an ELEXSYS Super High Sensitivity resonator operating at X-band (9.9 GHz). To avoid over modulation of the signal, magnetic field modulation amplitude was optimized at 0.1 or 0.2 mT with microwave power of 10 mW and frequency modulation of 100 kHz. The percentage yield of the reaction was calculated by double integration of the EPR signal recorded under non-saturating conditions and comparison with the doubly integrated signal of a MTSL standard sample of known concentration.

The line shapes of the EPR spectra of the MTSL label grafted on hdn_C175, hdn_n_C175 and hdn_C238 reveal important structural and dynamic features of Engrailed-2. The analysis of different Engrailed-2 constructs relies on the fact that the mobility of the label depends on the protein site onto which it is grafted.

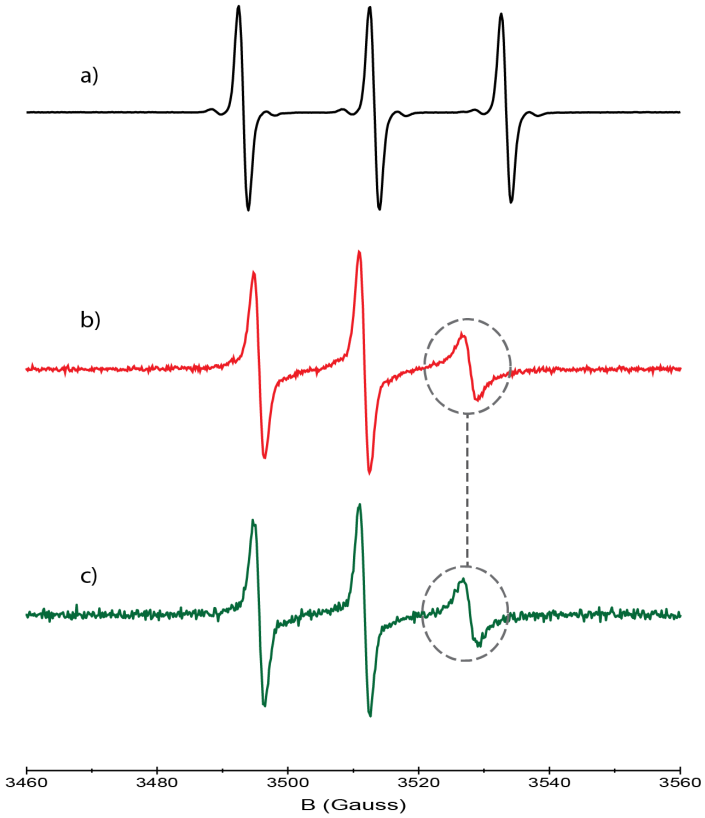


Figure 57 Experimental EPR spectra for free MTSL, hdn-C175 and hdn_n_C175. (a) The EPR spectrum of free MTSL is characteristic of free radical. EPR spectra of the radical grafted to two constructs of Engrailed 2: (b) hdn_C175 and (c) hdn_n_C175. The two spectra of the labeled protein show the same EPR line shapes.

The line shapes of EPR spectra of hdn_C175 and hdn_n_C175 are very similar (Figure 57). This shows that the dynamics of the hexapeptide are very similar in the presence and absence of the homeodomain. In both cases, a high mobility of the probe manifests a typical behavior of intrinsically disordered proteins.[108]

These observations suggest that the homeodomain does not appear to interact tightly with the hexapeptide in the disordered region, as the mobility of the MTSL radical in the hexapeptide of hdn_C175 and hdn_n_C175 does not change in the presence of homeodomain. This observation may seem at first contradictory to the PRE results, which indicate the presence of long-range contacts between the homeodomain and hexapeptide as discussed in PRE section.

However, these experimental observations can be reconciled in a series of potential scenarios:

1. The detection of a transient interaction between the hexapeptide and homeodomain would be difficult if the population of the bound form is small compared to the free form. This would be all the more possible as the signal-to-noise ratio in EPR spectra would not allow the observation of a minor form populated below ~ 10%
2. If the transient interaction is very fast on sub-nanosecond timescale, there should be no contribution to the line broadening recorded by the EPR line shape. The common overall tumbling motions could dominantly affect the spectra recorded at 9.9 GHz. In this case higher orientational resolution by using higher field can describe the precise orientation of the molecule.[109] However, this possibility can be excluded as the fastest possible free-bound reaction would be limited by segmental diffusion, which was shown to take place on time scales slightly slower than 1 ns.[110]
3. A fuzzy complex could be transiently formed. In a fuzzy interaction between hexapeptide and homeodomain, the disordered region would retain its disordered character.[111] However, we might expect the nanosecond time scale dynamics to be altered, even in a fuzzy complex.
4. The interaction of the hexapeptide and the homeodomain is very weak. The IDR would be constrained to a region close to the surface of the homeodomain, by electrostatic interactions, for instance. In that case, the observed PREs between the hexapeptide and the homeodomain would illustrate that the hexapeptide is most of the time within 15-20 Å of the homeodomain, rather than in a tight complex in a minor form.

The EPR spectrum of hdn_C238, where the spin label is grafted onto the loop region between the 2nd and 3rd helix of the homeodomain is shown in Figure 38. The comparison to the EPR spectra of nitroxide-labeled hdn_C175 and hdn_n_C175, show a broadening that is in agreement with the fact that the homeodomain tumbles slower than the hexapeptide. However, the difference of line width between the spectra of the

label grafted onto the hexapeptide or onto the homeodomain is not dramatic, so that only an important fraction of hexapeptide bound could be detected.

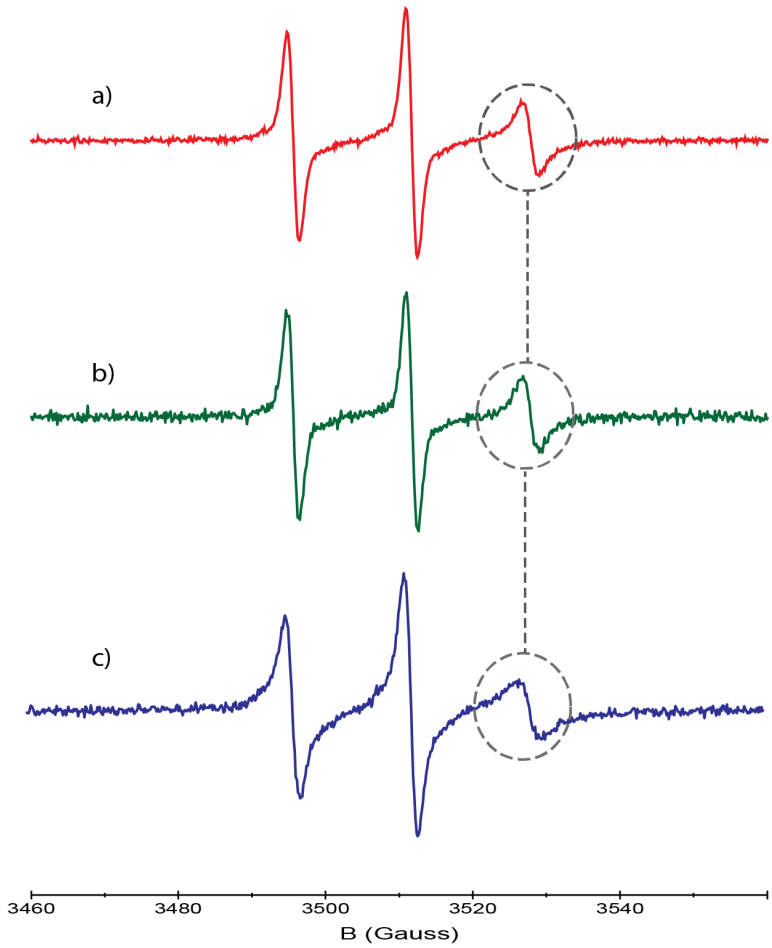


Figure 58: Comparison of the experimental EPR spectra of MTSL-labeled hdn-C175, hdn_n_C175 and hdn_C238. EPR spectra of the radicals grafted on (a) hdn_C175, (b) hdn_n_C175 and (c) hdn_C238.

Although no specific software (EasySpin) was used to simulate the EPR spectra, the motional regime that corresponds to the characteristic EPR line shapes measured on hdn_C175, hdn_n_C175 and hdn_C238 seem to belong to fast time scales (about 10^{-9} s) as mentioned in Figure 55. The fast motional regime results in line broadening due to existence of some degree of magnetic anisotropy, with no change in line positions. Clearly, fast time scale (10^{-9} s) motions of the label grafted to the homeodomain are

slower than those of the nitroxide label grated on the N–terminus of the hexapeptide (Figure 58).

We know from NMR relaxation and chemical shift analysis that the hexapeptide is significantly ordered in the 146-200 construct of Engrailed 2. Identical EPR spectra of hdn_C175 and hdn_n_C175 show that the hexapeptide region does not loose its structure in the absence of the homeodomain and that it is not tightly bound to the homeodomain in a significant fraction of the time. The two most likely hypotheses that are in agreement with both EPR and PRE data are that the hexapeptide would be bound to the homeodomain a small fraction of the time or that the hexapeptide does not form any tight complex with the homeodomain but remains most of the time at a medium distance of the homeodomain. In both cases, the presence of the IDR (including the hexapeptide) would decrease slightly the affinity of the homeodomain for DNA, which is what was observed by our fluorescence anisotropy binding studies.

Chapter 6 Residual Dipolar Couplings (RDCs) to map out the conformational space of disordered region of Engrailed-2

Chapter 6 Residual Dipolar Couplings (RDCs) to map out the conformational space of the disordered region of Engrailed-2

6.1 Principle of Residual Dipolar Couplings (RDCs)

In solution NMR, under isotropic conditions, dipolar couplings between two nuclear spins in a molecule average to zero due to fast tumbling of the molecule. This averaging is at the origin of the high-resolution nature of solution NMR spectra. By contrast, these dipolar couplings are prominent in solid-state NMR. Dipolar interactions are averaged out by mechanically spinning the sample at the magic angle (54.7° with the B_0 field) to obtain sharp resonances in NMR spectrum.[112]

In solution, residual dipolar couplings (RDCs) can be induced by partially aligning the molecule (dissolved in a slightly anisotropic solution) relative to the magnetic field leading to a slight bias in the distribution of protein orientations. The partial alignment of the molecule leads to residual dipolar couplings (RDCs), which can be measured as perturbations to scalar couplings. These RDCs depend on the residual orientation of each internuclear vector.[27] The residual orientation of all internuclear vectors in a protein can be related to their orientation in the molecular frame (*i.e.* the structure of the protein) and the overall alignment of the protein in the anisotropic solution. RDCs are thus a very useful source of structural constraints. Note that a simple cosine dependence on a single angle provides ambiguous information and alignment in several different media is necessary to use RDCs as the only source of structural information. In proteins, ^{1}H-^{1}H NOEs can generally be detected only for protons within 5 Å. Thus, NOEs provide only local constraints and have limited information content for the coherence of a structure over longer distances. Unlike NOEs, the information on the orientations that are distant in the structure contained in dipolar couplings is relative to the magnetic field and hence doesn't depend on distance. Thus RDCs provide valuable global orientational restraints that add wealth

of information to the strictly local NOEs.[28] The most common use of RDCs is to refine and validate the structure determined by X-ray crystallography or solution NMR. The determined structure coordinates can be readily used to fit RDC data to the corresponding bond vector directions within the molecular frame.[113] Moreover, the RDCs identify precisely the long-range order in structured proteins and is sensitive to transient long-range order in IDPs.[114,115] Due to their sensitivity to internuclear bond-vectors, RDCs have been efficiently used to describe the local conformational behavior of IDPs.[116–120]

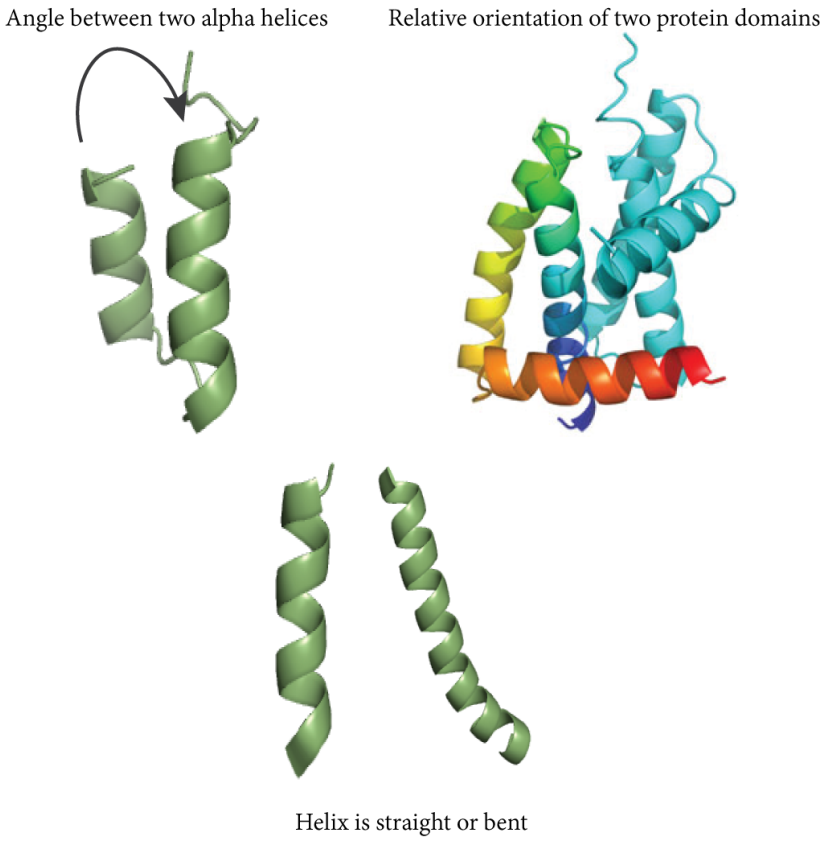


Figure 59: The RDCs contain valuable information of orientation of the bond vector relative to the external magnetic field, which provides orientational short- and long-range constraints in protein structure determination.

The dipolar couplings of two spins observed in partially aligned medium can be expressed as,[121]

$$D_{IS}(\theta) = \frac{\hbar\gamma I\gamma S}{4\pi^2 r^3 IS}[1 - 3cos^2\theta] \quad (23)$$

where $\hbar = \frac{h}{2\pi}$, γ_I and γ_S are the gyromagnetic ratios of spin I and spin S respectively, r_{IS} is the internuclear distance and θ is the average angle between the bond vector and the external magnetic field.

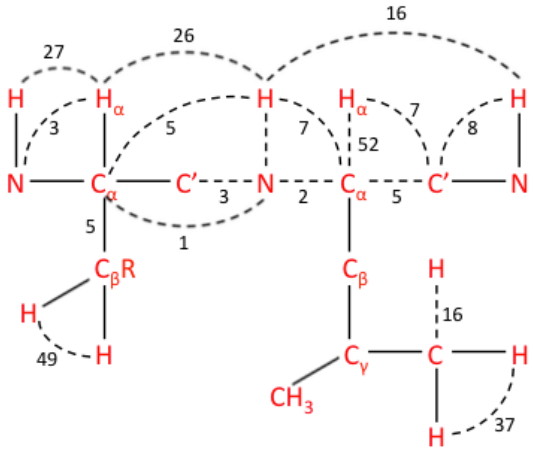


Figure 60: RDCs that can be measured experimentally using various alignment media. Here figure shows the calculated maximal RDCs in Hz (relative to a maximal RDC of 25 Hz of a ^{1}H-^{15}N spin pair) on proteins. Adapted from Ref.122.[122]

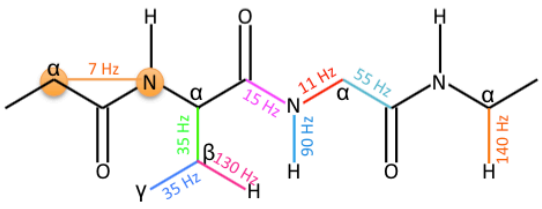
6.2 Selection of alignment media

The partial alignment of molecules can be achieved by using anisotropic media. Different proteins will be stable in a different subset of alignment media and the availability of RDCs with more that one overall alignment tensor is necessary to reduce the ambiguity in the orientation information. Thus, many different alignment media have been designed to vary both the overall orientation of the induced alignment and the physical and chemical conditions of use. In the selection of alignment media, care must be taken to avoid unwanted interactions between molecule and alignment media that can change the correlation time of the molecule. The compatibility between pH, ionic strength, the temperature of the sample, and the alignment media are crucial criteria for the selection of alignment media.[123] The most

commonly used alignment media are liquid crystalline phases, bicelles, Pf1 phages and compressed or stretched polyacrylamide gels. The first application of external alignment media (following the pioneering work of Tolman et al. on myoglobin[124]) to a protein (ubiquitin) was done in bicelles as it produces primarily steric alignment.[27] The charge at the surface of bicelles can be tuned by the introduction of lipid group heads with positive or negative charge in order to prevent the association between the protein and the alignment medium to preserve high-resolution spectra. Preparation of alignment samples with Pf1 phage requires simply the dilution of the protein sample in a solution of phage. Note that Pf1 phages are negatively charged and not suitable for proteins far below their isoelectric point. Preparation of partially aligned samples with polyacrylamide gel requires overnight polymerization of either positively or negatively charged polyacrylamide gel in a suitable buffer followed by gel drying. The dried gel is then soaked by the protein solution.[125] The gels can be either compressed or stretched in adapted NMR tubes. The latter option provides the most uniform alignment.

6.3 The NMR techniques for measuring RDCs

RDCs are measured as perturbations to scalar couplings. The basic one-bond coupling constants that can be measured by using different pulse programs are summarized in Figure 61:



1 bond couplings	
$^{1}J_{C\text{-}H}$	140 Hz
$^{1}J_{N\text{-}H}$	90-100 Hz
$^{1}J_{N\text{-}^{13}C=O}$	15 Hz
$^{1}J_{N\text{-}^{13}C\alpha}$	11 Hz
$^{1}J_{C\alpha\text{-}C=O}$	55 Hz
$^{1}J_{C\alpha\text{-}C\beta}$	35 Hz

Figure 61: The basic one-bond coupling constants that can be measured in different alignment media to calculate the RDC effect. The magnitude of these couplings depends on the proximity and nuclear magnetic moments of the coupling partners.

A large number of NMR experiments have been developed to accurately measure the coupling constants between two nuclei. There are two approaches to measure RDCs: *frequency resolved methods* that are based on the direct measurement of the frequencies of the multiplet components and *J-modulated methods,* where the evolution under couplings during a variable delay leads to a modulation of the signal amplitude. In both cases, RDCs are determined from the difference between the measured scalar coupling constant in isotropic phase and the effective coupling constant in an aligned medium.[122] There are several *frequency resolved* and *J-modulated* 2D and 3D experiments that are frequently used to measure *J*-couplings and RDCs.

6.4 IPAP (frequency resolved 2D based approach)

The in-phase/anti-phase (IPAP) is practically the simplest method to measuring ^{1}J couplings. Two interleaved two-dimensional experiments are recorded with no decoupling in the indirect dimension. One experiment displays the evolution of the in-phase (IP) doublet and the other one the anti-phase (AP) doublet.

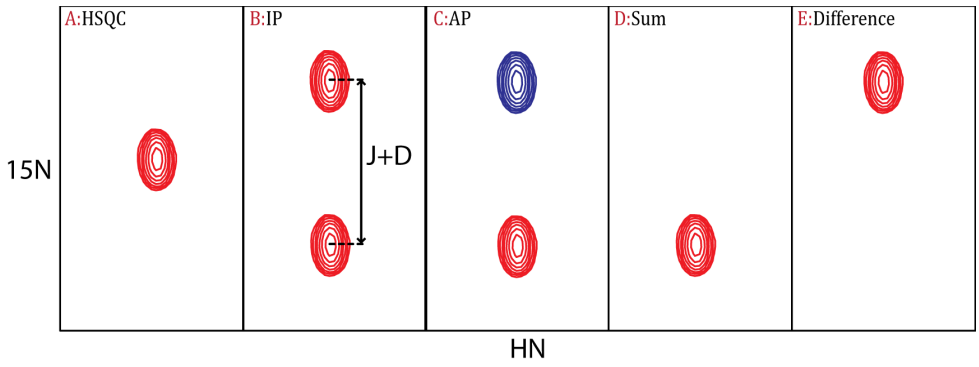


Figure 62: Basic principle of IPAP experiment. Two interleaved experiments that correspond to upfield and downfield components. Couplings are simply extracted by direct measurement of peak splitting (Hz) between these two interleaved spectra. A: no splitting (normal decoupled HSQC); B: In-Phase doublet, C: Anti-Phase doublet, D & E, the sum and the difference of B & C. In anisotropic phase J+D-splitting, the sum of the scalar coupling and residual dipolar coupling is measured.

The sum and difference of the two resulting spectra correspond to spectra that display only one of the two components of the doublet appear. The frequencies can be

measured on each spectrum and the difference is the effective coupling. This method is widely used to measure accurately large NH (^{1}H-^{15}N) and CH ($^{13}C_{\alpha}$-$^{1}H_{\alpha}$) residual dipolar couplings.[126] The IPAP approach can also be easily implemented in higher-dimensionality experiments (most often 3D).[127,128]

6.4.1 J-modulated (intensity based Method)

This method involves the measurement of NMR signal intensities that are modulated by an evolution under the coupling. At least two experiments are recorded. Couplings are not allowed to evolve in the reference experiment. In the second experiment the signal is modulated according to a cosine function and without modulation are recorded. The couplings are then derived from intensity ratio of two experiments. Alternatively, a series of experiments can be recorded, with variable delays for the evolution under couplings. Precise values of couplings can then be fitted to the modulated amplitude. This method can be used to extract smaller couplings (^{15}N-$^{13}C'$, $^{1}H^{N}$-$^{13}C'$) where the direct measurement of peak splitting becomes difficult in a coupled correlation (IP spectrum) because of severe spectral overlap.[129,130]

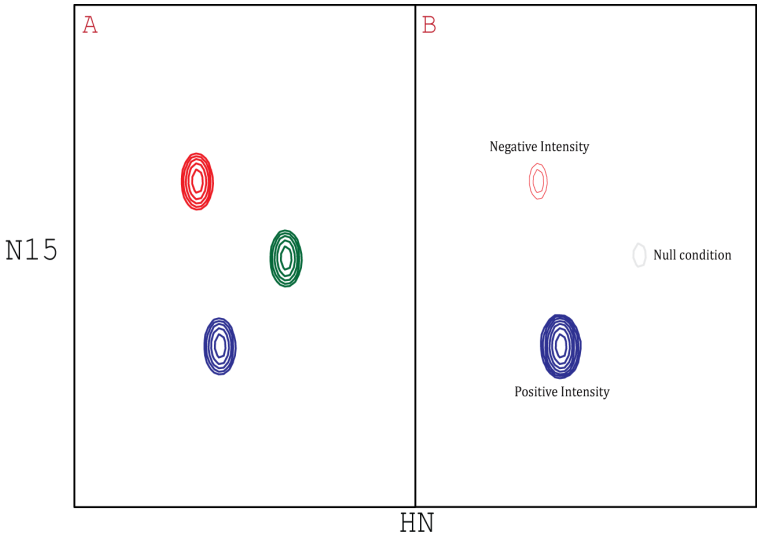


Figure 63: Basic concept of intensity based method to determine the RDCs. The intensity ratio of two interleaved experiments that are modulated by coupling evolution provides coupling constants in isotropic and aligned sample conditions. (A) Reference spectrum, (B) Attenuated spectrum, Positive intensity indicates J + JD are positive; negative intensity indicates J + JD are negative; loss of peak indicates a J + JD value close to the null condition.

In order to acquire a sufficient ensemble of experimental data to refine quantitatively the conformational space of the disordered region of Engrailed-2, we recorded a series of experiments to collect several sets of one-bond residual dipolar couplings. Engrailed-2 interacts with bicelles, which alters its conformation[31]. Besides, Engrailed-2 is basic (pI =10.6) and can interact with phages. Thus, we decided to use a stretched polyacrylamide gel to align Engrailed 2 and measure residual dipolar couplings.

6.5 Procedure for casting polyacrylamide gel

To avoid any change in rotational correlation times and too large of an alignment, it is advisable to prepare gel with concentration in the range of 3-8% (w/v). The higher concentration that gives larger RDCs causes a lines broadening and lower signal to noise ratio. The larger diameter of the gel gives larger RDCs but instability of the gel becomes higher with larger diameter.

To cast a 5.4 mm, 4% (w/v) acrylamide gel, following ingredients and procedure have been used.

Ingredients	**Quantity**
Acrylamide 30% (37.5:1, acrylamide/bis-acrylamide)	200 µL
TEMED (Tetramethylethylenediamine)	5 µl
APS (Ammonium Persulfate)	5 µl
Buffer sodium phosphate (200mM, pH=7)	1290 µL
Total volume	1.5 ml

Table 4: List and quantity of ingredients used to cast 4% (w/v) acrylamide gel to record one-bond couplings in Engrailed-2.

All ingredients except TEMED were mixed in an Eppendorf tube and the mixture was filtered through a 0.2 μm pore size filter to remove any polymerized impurities. The mixture was then degassed to remove air bubbles that can induce perforation in the gel polymerization process. Finally TEMED was added and all ingredients were thoroughly mixed by vortex. The mixture was immediately transferred into a 5.4 mm diameter piston with a Teflon plug at the bottom and covered with parafilm at the top so that polymerization occurs in an airtight environment (evaporation of liquid would lead to undesired drying of the gel). The mixture was allowed to polymerize overnight at room temperature.

The gel was then pushed out of the piston with a rod onto aluminum foil. The gel was cut to desired size and immersed in sodium phosphate buffer (200 mM, pH = 7) inside a beaker. The gel was rinsed in milli-Q water with gentle shaking for 3-4 hours. This operation was repeated three times. For the last rinsing step, the gel was left in milli-Q water overnight with gentle shaking. The rinsed gel was dried in a Petri glass dish and then transferred in an eppendorf tube and immersed into 275 μl of 0.5 mM uniformly ^{15}N-^{13}C labeled sample of Engrailed-2 dissolved in 40 mM succinate buffer (pH = 6.0) with 10% D_2O. The gel and protein sample were kept in the eppendorf tube for 48 - 72 hours with continuous shaking to completely homogenize the sample in the gel. The gel was then introduced into an NMR tube by carefully pushing the Teflon bottom of the piston to push the gel into the NMR tube. The sample tube was then introduced in the spectrometer to verify the local anisotropy by recording the deuterium splitting of the solvent signal. The sample in the tube was left at room temperature for another 24 - 48 hours before recording the NMR experiment to measure RDCs. The isotropic ^{15}N-^{1}H scalar couplings were measured with similar sample concentration in the similar buffer as used for the gel sample.

6.6 Results and Discussion

Preliminary experiments were recorded on a 500 MHz spectrometer equipped with a cryoprobe to verify the degree of alignment. The water deuterium splitting was found to be 6 Hz. No evaluation of the deuterium splitting was found over the course of

preliminary experiments, which suggested that the gel was stable over a few days. A simple, sensitivity-enhanced HSQC-edited IPAP experiment was recorded, which separates the two components of the ^{15}N doublet in the indirect dimension to measure the N-H couplings.[131] The couplings are simply extracted by direct measurement of peak positions (Hz) in the nitrogen dimension. The maximum one-bond ^{1}H-^{15}N residual dipolar couplings with 4% acrylamide gel sample were found to be of the order of ± 4 Hz, which was clearly insufficient. Gels with higher concentration of acrylamide were cast to achieve approximately the desired ± 20 Hz couplings. Similar protocol and experimental setup were used to analyze the one-bond ^{1}H-^{15}N residual dipolar couplings for 5% and 6% (5.4 and 6 mm diameter respectively) of acrylamide gel that yielded couplings ranging from ~ – 6 Hz to ~ 8 Hz. The maximum ^{1}H-^{15}N residual dipolar couplings (ranging from ~ – 18 Hz to ~ 18 Hz) were measured with a 7% acrylamide gel (6 mm diameter gel). The constant deuterium splitting showed that the sample was stable over the period of time necessary to record sets of experiments to measure various one-bond residual dipolar couplings.

Final experiments were recorded on a (7% w/v) acrylamide gel (6 mm diameter) containing 0.5 mM uniformly ^{15}N-^{13}C labeled sample of Engrailed-2 dissolved in 40 mM succinate buffer (pH = 6.0) with 10% D_2O. The one-bond ^{1}H-^{15}N, C_α-H_α and C_α-C' residual dipolar couplings were recorded on a 800 MHz spectrometer using following parameters:

Couplings	Pulseprogram (Bruker ID)	Complex Data Points $^{1}H \times ^{15}N \times ^{13}C$
^{1}H-^{15}N	HNCO based[127] (hncogprc3d6)	1792×70×96
C_α-H_α	HNCO based[128] (hncogprc3d5)	1792×48×80
C_α-C'	HNCO based[127] (hncogprc3d1)	1792×48×120

Table 5: Summary of experimental parameters and Pulseprogram (Bruker ID) used to measure one-bond ^{1}H-^{15}N, C_α-H_α and C_α-C' residual dipolar couplings in Engrailed-2.

6.6.1 ^{1}H-^{15}N-dipolar coupling

The ^{1}H-^{15}N residual dipolar couplings are easy to measure and have been used most frequently in many studies.

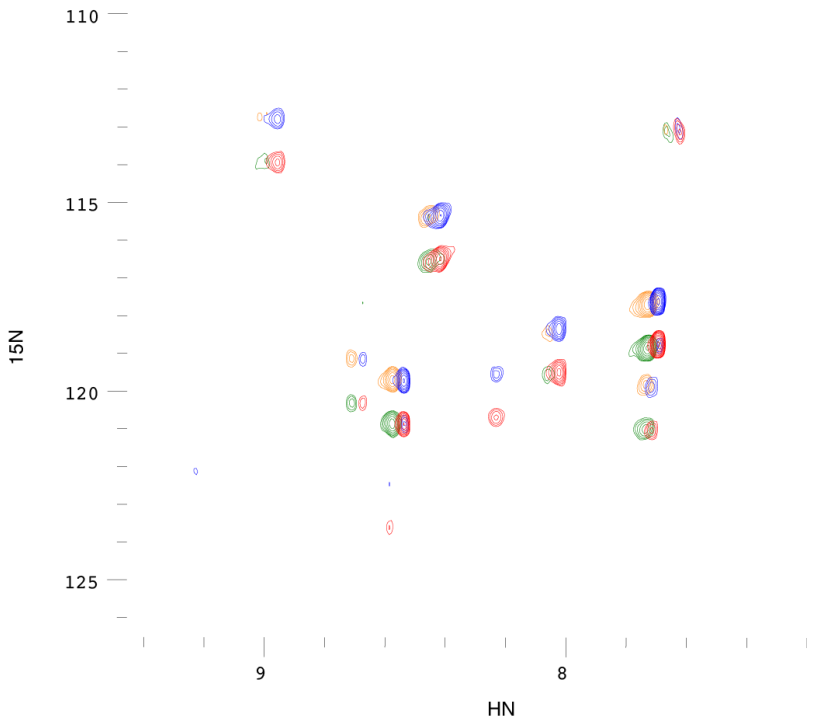


Figure 64: ^{1}H-^{15}N plane (CO chemical shift 177.02 ppm) from the IPAP HNCO spectra that correspond to low-frequency and high-frequency components in isotropic solution (blue & red, respectively) and in the aligned (orange & green respectively) state of Engrailed-2. The separation of the two peaks in the vertical dimension is the sum of the RDC and the scalar coupling (J+D) in the aligned spectrum and the scalar coupling alone (J) in the isotropic spectrum.

One-bond ^{1}H-^{15}N couplings were measured using sensitivity-enhanced HNCO-based IPAP three-dimensional experiments. Couplings were measured from peak positions in the ^{15}N dimension[127] to investigate the conformational space of discorded region of Engrailed-2. The one-bond ^{1}H-^{15}N residual dipolar couplings for Engrailed-2 were calculated as the difference of splittings recorded in the gel sample (aligned state) and in the solution sample (isotropic state).

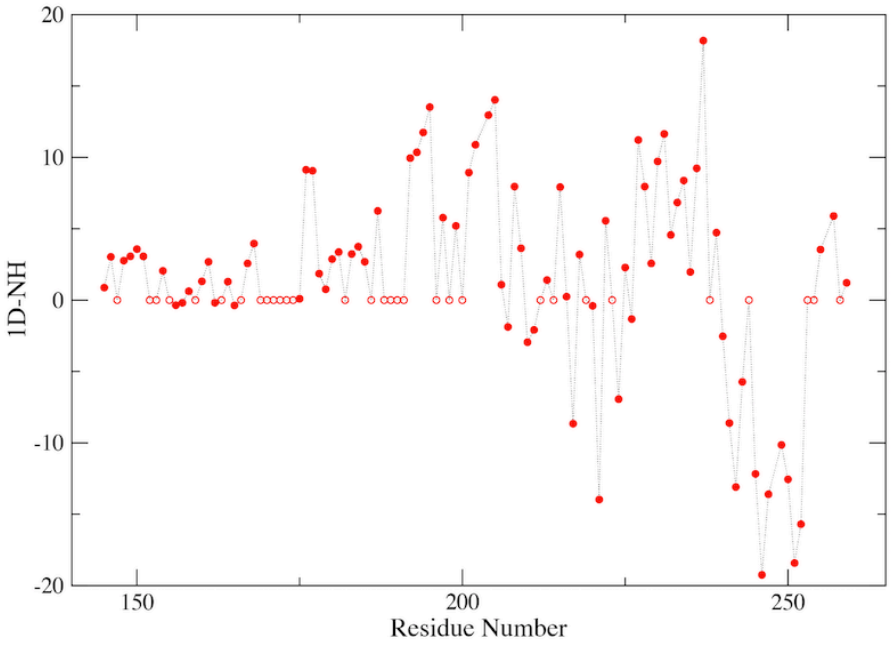


Figure 65: The measured values of one-bond *^{1}H-^{15}N* residual dipolar couplings for Engrailed-2 calculated as the difference of doublet splittings recorded in the aligned (gel sample) and isotropic state (solution sample). The dipolar coupling values displayed as a function of residue number for Engrailed-2. Open circles represent the missing residues in the aligned state as corresponding peaks were too broad to appear in the aligned state.

The one-bond ^{1}H-^{15}N residual dipolar couplings assigned for Engrailed-2 were in the range of ~ – 19 Hz to ~ 19 Hz in the homeodomain and from 0 to 13 Hz in the disordered N-terminal part (note that RDCs could not be determined in the hexapeptide).

6.6.2 $^{13}C_{\alpha}$-$^{1}H_{\alpha}$-dipolar couplings

The $^{13}C_{\alpha}$-$^{1}H_{\alpha}$-dipolar couplings are almost twice larger (up to 50 Hz) than ^{1}H-^{15}N residual dipolar couplings. In structure determination and refinement, these couplings contained better orientational information compared to ^{1}H-^{15}N residual dipolar couplings.[122] Scalar and dipolar one-bond $^{13}C_{\alpha}$-$^{1}H_{\alpha}$-dipolar couplings were measured using sensitivity-enhanced HNCO-based 3D experiments where the one-bond $^{13}C_{\alpha}$-$^{1}H_{\alpha}$ couplings can be measured in the carbonyl dimension.[128] The one-bond $^{13}C_{\alpha}$-$^{1}H_{\alpha}$-dipolar couplings for Engrailed-2 were calculated as the difference of doublet

splittings recorded in the gel sample (aligned state) and in the isotropic solution sample.

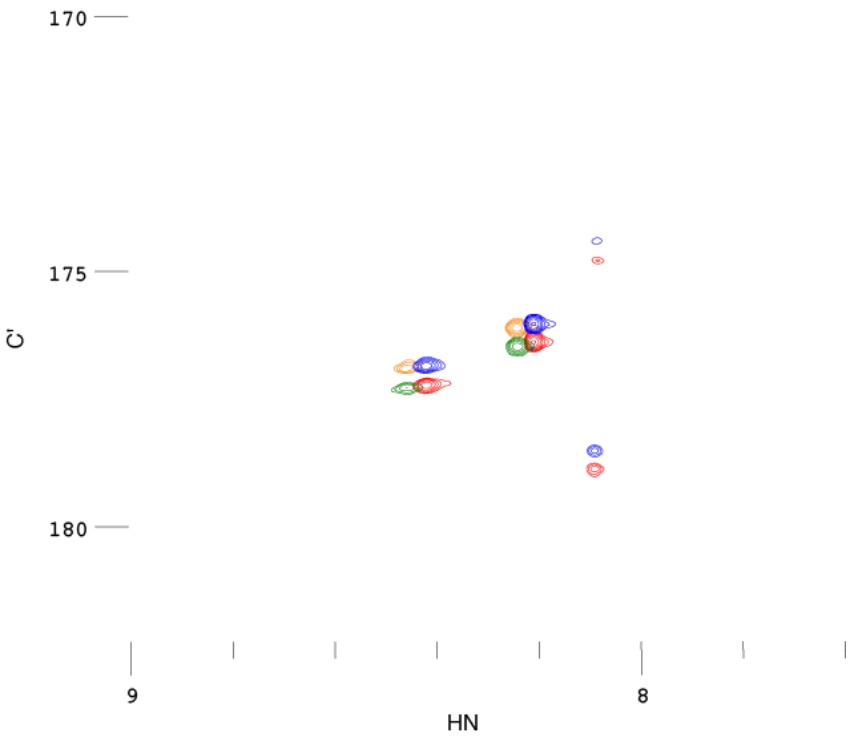


Figure 66: ^{1}H-^{13}CO plane (^{15}N chemical shift 115.87 ppm) of the experiments that correspond to low-frequency and high-frequency components of the apparent ^{13}CO doublet in isotropic solution (blue & red) and in the aligned state(orange & green) of Engrailed-2. The separation of the two peaks in the vertical dimension is the sum of the RDC and the scalar coupling (J+D) in the aligned spectrum and the scalar coupling alone (J) in the isotropic spectrum.

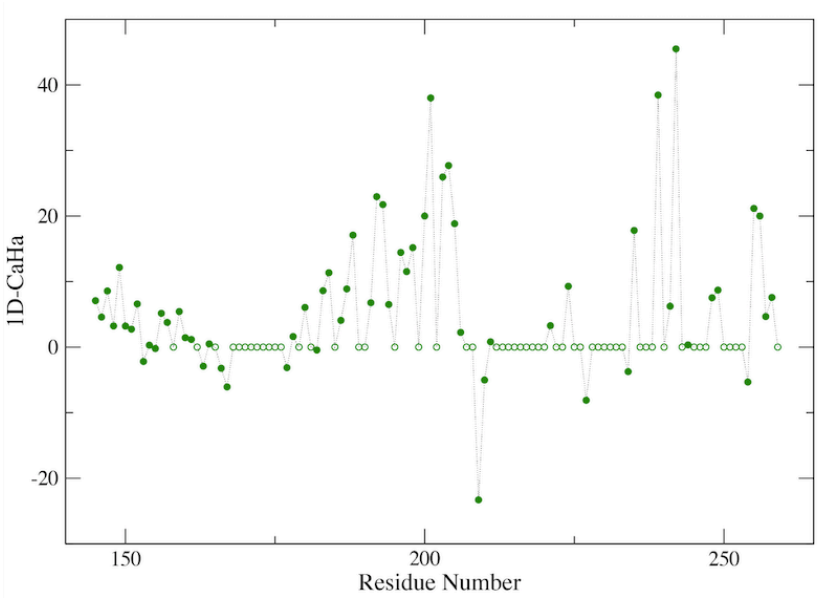


Figure 67: The measured values of one-bond $^{13}C_{\alpha}$-H_{α} residual dipolar couplings for Engrailed-2. The residual dipolar coupling values are displayed as a function of the residue number for Engrailed-2. Open circles represent the missing data in the aligned state as corresponding peaks were too broadened for frequencies to be measured in the aligned state.

The values of one-bond $^{13}C\alpha$-$^{1}H\alpha$ dipolar coupling for Engrailed-2 were in the range of ~ – 24 Hz to ~ 45 Hz in the homeodomain and between – 6 Hz and 22 Hz in the disordered region (note that, again, no RDC could be measured in the hexapeptide).

6.6.3 $^{13}C_{\alpha}$-C′-dipolar couplings

The $^{13}C_{\alpha}$-C′-dipolar couplings are typically about one-fifth of ^{1}H-^{15}N residual dipolar couplings. These couplings can be easily measured with 3D IPAP-HNCO experiments with no decoupling on $^{13}C_{\alpha}$ during the C′ chemical shift evolution.[127]

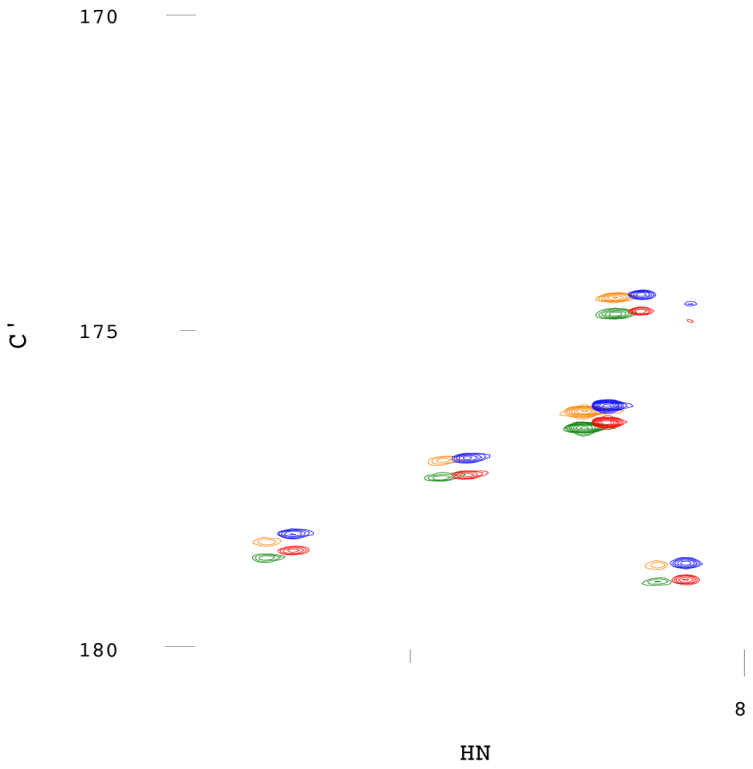


Figure 68: ^{1}H-^{13}CO plane (^{15}N chemical shift 115.72 ppm) of the two spectra that correspond to the low-frequency and high-frequency components of the ^{13}CO doublet in isotropic solution (blue & red) and in the aligned state (orange & green) of Engrailed-2.

The values of one-bond $^{13}C_{\alpha}$-C′-dipolar couplings for Engrailed-2 were in the range of ~ – 10 Hz to 9 Hz in the homeodomain and – 5 Hz to 2 Hz in the disordered region.

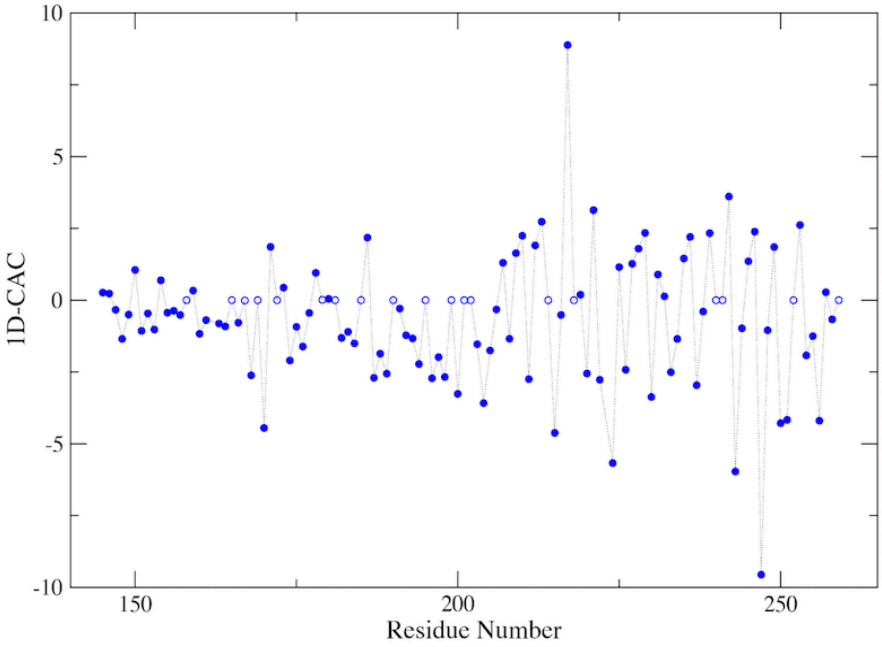


Figure 69: One-bond $^{13}C_{\alpha}$-C′-residual dipolar couplings for Engrailed-2 displayed as a function of residue number for Engrailed-2. Open circles represent the missing residues in aligned state as corresponding peaks were too broadened for frequencies to be measured in the aligned state.

The one-bond ^{1}H-^{15}N, C_{α}-H_{α} and C_{α}-C′ residual dipolar couplings measured for Engrailed-2 show a distribution of RDCs along the sequence. The one-bond RDC values measured in the homeodomain of Engrailed-2 indicate a rather high degree of alignment. The lower values of RDCs measured in the disordered region illustrates the conformational heterogeneity as averaging of multiple diffusion tensors of the disordered region.[27] The distribution of ^{1}H-^{15}N RDCs along the sequence of the homeodomain shows a pattern that follows closely the secondary structure. RDCs in the 2nd helix of the homeodomain are positive while 1st and 3rd helix shows partially and fully negative values, respectively. This observation somehow suggests that all three helices adopt different orientation with respect to the magnetic field.[33,34] Dipolar waves can be observed clearly in helices 2 and 3 in particular. Note that dipolar waves that characterize recognizable signatures for helical elements[33] can be observed at least for helices 2 and 3 of the homeodomain.

As mentioned earlier in chapter 2, the solution structure of the homeodomain of Engrailed-2 has been recently solved by NMR. We used the NMRPipe program DC to fit the alignment tensor to the measured ^{1}H-^{15}N dipolar couplings from the three helices of the homeodomain (excluding RDC values for loops and the disordered region). The Order Matrix Fitting method is used by DC. This method accounts no prior assumption about the alignment tensor parameters and performs the fit by linear least squares analysis. This analysis gives a set of best-fit alignment tensor parameters and also generates a collection of calculated dipolar couplings, which can be compared to the measured values. The comparison of the measured and calculated RDCs for the homeodomain of the Engrailed-2 is described in Figure 70.

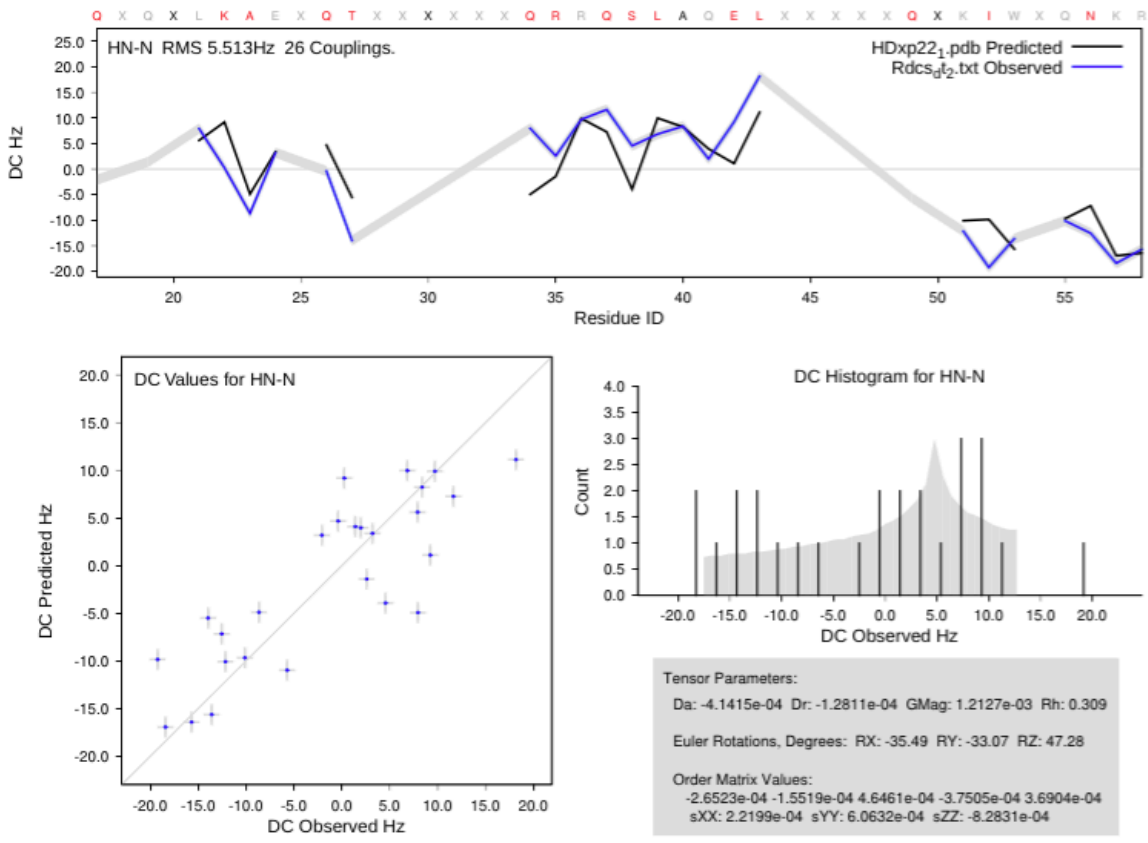


Figure 70: Comparison of the measured and calculated ^{1}H-^{15}N RDCs for residues located in the helices of the homeodomain of Engrailed-2

The comparison shows a rather disappointing correlation between calculated and measured values but there is no systematic deviation. The difference between the

measured and calculated RDCs can arise from problems in the solution structure (which was not refined using RDCs) or possibly from the fluctuations of the alignment tensor due to the dynamics of the disordered region of the protein.

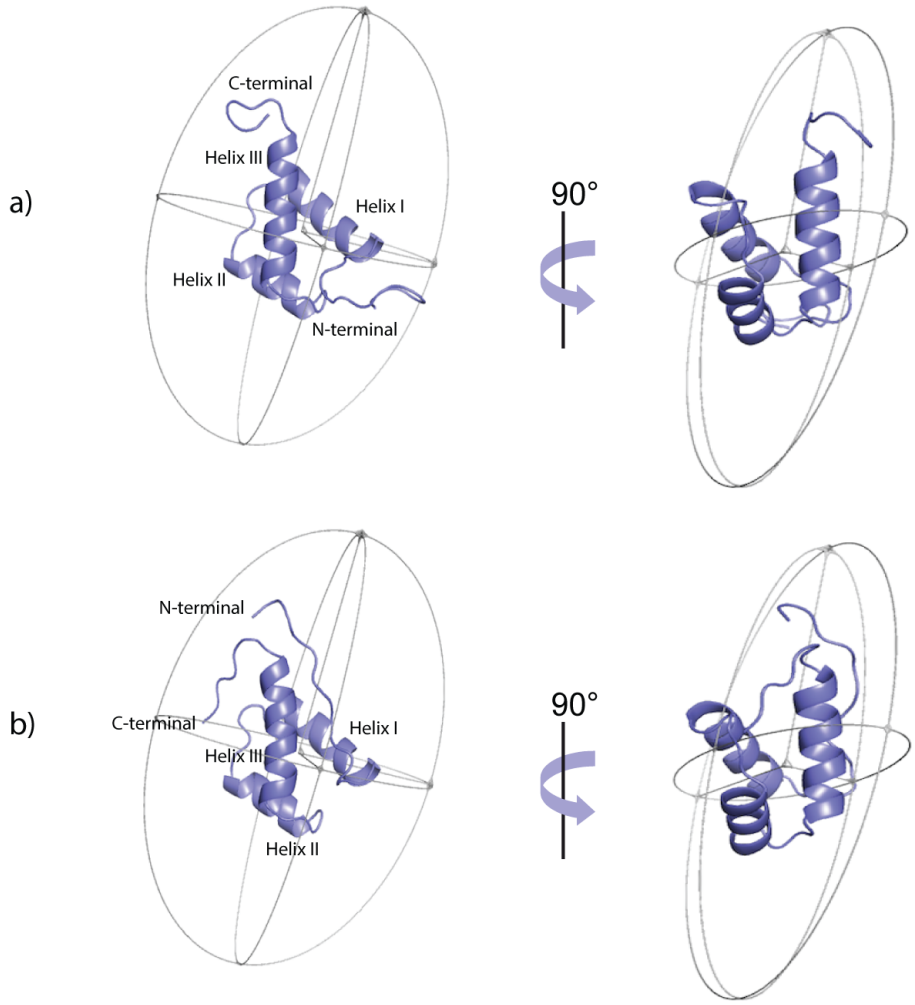


Figure 71: Graphical representation of the fitted alignment tensor in the molecular frame of two conformers (a) and (b) of the homeodomain.

The alignment tensor of the homeodomain is represented in Figure 71. We see that helices 1 & 2 are aligned transverse while the third helix is positioned along the main component of the alignment tensor. The shape of the alignment tensor provides some insights on the volumes where the IDR is most likely to be located within the conformational space. The alignment medium proceeds mostly from the steric interaction between the protein and the gel. The two sides of the protein that face the smallest component of the alignment tensor are thus probably least likely to be

covered significantly by the IDR. On the other hand, the alignment tensor of the homeodomain is insufficient to provide more information on the conformational space of the IDR.

The values of ^{1}H-^{15}N RDCs in the disordered region are variable but mostly positive, in contrast to what would be expected in a fully disordered polypeptide[116]. Interestingly, all RDCs are significantly different in the IDR at the N-terminus of the hexapeptide and the IDR between the hexapeptide and the homeodomain. Importantly, the typical amplitude of RDCs is larger for all three sets (^{1}H-^{15}N, $^{1}H_{\alpha}$-$^{13}C_{\alpha}$, and $^{13}C_{\alpha}$-$^{13}C'$) in the IDR between the homeodomain and the hexapeptide, which indicates a higher level of order. This observation is somehow in agreement with the IMPACT analysis, which shows a higher prevalence of motions slower than 1 ns in this IDR compared to the IDR at the N-terminus. Unfortunately, almost no information is available about the hexapeptide as peaks of this region were missing in most spectra.

$^{13}C_{\alpha}$-$^{13}C'$ residual dipolar couplings are small, overall about one-fifth of the ^{1}H-^{15}N residual dipolar couplings. The variation of orientations of $^{13}C_{\alpha}$-$^{13}C'$ vectors throughout helices is illustrated by the large variations of RDCs along the sequence of the homeodomain. The $^{13}C_{\alpha}$-$^{13}C'$ residual dipolar couplings were almost complete in both disordered and folded regions of engrailed-2. In particular, this RDC dataset is almost complete at the hexapeptide site.

Conclusive Discussion

Conclusive Discussion

IDPs are a class of proteins that do not follow the classical structure/function paradigm. These proteins lack a stable 3D structure but yet play vital biological functions. Homeotic Complex (HOX) homeoproteins constitute a large family of transcription factors that regulate the expression of specific target genes. A HOX protein contains a DNA binding homeodomain that folds in a tri-helical core and mostly bears disordered C and N-terminal tails. These homeoproteins are believed to be DNA binders with low specificity and require other partner proteins to attain high DNA binding specificity. These proteins attain DNA binding specificity through the interaction of their disordered region with the partner proteins within ternary complexes.

These findings lead to set the major goals of our study to investigate the role of disordered region of the homeoprotein Engrailed-2 with the aim to

- Explore the role of the disordered region in DNA binding in the absence of partner protein
- Characterize the conformational space of the Engrailed-2 especially the disordered region and the exploration of its dynamics

To answer these questions we carried out DNA binding studies using fluorescence anisotropy. Several constructs of Engrailed-2 including the homeodomain and various lengths of the disordered region were studied. The measured dissociation constants indicate that the homeodomain binds with strong affinity to the DNA. This affinity is further enhanced in a construct that includes the homeodomain and a series of basic residues at the N-terminal extension of the homeodomain. This enhancement most likely comes through the interaction of these basic residues with the phosphate backbone of the DNA. Interestingly, this gain in affinity was lost when the protein contained a longer part of the disordered region at the N-terminus of the homeodomain. The long disordered region in Engrailed thus decreases the affinity for DNA. This suggests that the disordered region is involved in an auto-inhibition

mechanism and thus contributes to the specificity of DNA binding in the presence of other partner transcription factors both positively by increasing affinity through the interaction with the binding partner and negatively, by decreasing affinity in the absence of a binding partner. A hypothetical mechanism of auto-inhibition would be the shielding of the DNA binding sites by the disordered region. The possible proximity of the DNA binding sites to the disordered region was explored with paramagnetic relaxation enhancements (PRE), which is the prime method to investigate long-range contacts by NMR. The long-range contacts identified between the disordered region and the homeodomain show that the IDR occupies a volume that covers parts of the homeodomain in the conformational ensemble of Engrailed-2. The disordered region was found to be in contact with both the third helix of the homeodomain, which fits into the major groove of DNA, and the tail region at the beginning of the homeodomain, which fits into the minor groove.

The disordered region of Engrailed-2 comprises a small fragment (called hexapeptide, residues 169-174), which is the protein-protein interaction site in ternary complexes with DNA. Chemical shifts show that the hexapeptide has a higher structural propensity compared to the rest of the disordered region. We used EPR to test whether the observed long-range contacts were driven by a direct interaction between the hexapeptide and the homeodomain. EPR is a sensitive method to identify changes in dynamics due to protein-protein or protein-ligand interaction. EPR spectra show that the hexapeptide region does not loose its structure in the absence of the homeodomain and is not tightly bound to the homeodomain. The interaction of the IDR with the homeodomain is weak and possibly driven by electrostatic interactions between the acidic residues at the N-terminus of the IDR and the basic residues at the C-terminus of helix three or at the N-terminus of the homeodomain.

A proper understanding of the auto-inhibition mechanism thus requires to better define the conformational space of Engrailed-2. Additional experimental constraints can be obtained with the measurement of residual dipolar couplings (RDCs) in anisotropic media. RDCs provide valuable global orientational restraints and have been efficiently used to describe the conformational ensemble of the IDPs. RDCs were significantly lower in the disordered region at the N-terminus of the hexapeptide and

the disordered region between the hexapeptide and the homeodomain. The large amplitude of RDCs in the disordered region between the homeodomain and the hexapeptide likely indicates a higher level of order. The alignment tensor of the homeodomain provides insights on the overall localization of the IDR. The calculation of the conformational ensemble using ASTEROIDS (work in progress in the group of Martin Blackledge) will help to further refine the conformational space of the Engrailed-2.

In addition, we have recorded a broad set of relaxation rates and introduced a new approach for the Interpretation of Motions by a Projection onto an Array of Correlation Times (IMPACT) to analyze multiple-field relaxation data in disordered proteins. IMPACT significantly increases the number of correlation times by defining an array of fixed correlation times and thus explores the most relevant correlation times for backbone motions (pico-nano second) in the disordered region of Engrailed-2. Interestingly we confirm that the hexapeptide is partially ordered. We also find a higher contribution of the supra nanosecond motions in the disordered region between the homeodomain and the hexapeptide than in the disordered region at the N-terminus of the hexapeptide, in accordance with RDCs observations.

In conclusion:

- We confirm the role of the disordered region in the modulation of the affinity of the Engrailed-2 to DNA, including weak auto-inhibition.
- We show that the disordered region partially shields the interaction surface of the homeodomain with DNA.
- We have recorded a large set of experimental data to refine the conformational space of the Engrailed-2 and to understand the auto-inhibition mechanism.
- We have also shown, with a new approach, the diversity of ps-ns motions in disordered region of Engrailed-2

References

(1) Mirsky, A. E.; Pauling, L. *Proc. Natl. Acad. Sci. U. S. A.* **1936**, *22*, 439.
(2) Erickson, J. O.; Neurath, H. *J. Gen. Physiol.* **1945**, *28*, 421.
(3) Uversky, V. N.; Ptitsyn, O. B. *Biochemistry (Mosc.)* **1994**, *33*, 2782.
(4) Ptitsyn, O. B. *Trends Biochem. Sci.* **1995**, *20*, 376.
(5) Bode, W.; Schwager, P.; Huber, R. *J. Mol. Biol.* **1978**, *118*, 99.
(6) Williams, R. M.; Obradovi, Z.; Mathura, V.; Braun, W.; Garner, E. C.; Young, J.; Takayama, S.; Brown, C. J.; Dunker, A. K. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* **2001**, 89.
(7) Tompa, P. *Trends Biochem. Sci.* **2002**, *27*, 527.
(8) Galea, C. A.; Wang, Y.; Sivakolundu, S. G.; Kriwacki, R. W. *Biochemistry (Mosc.)* **2008**, *47*, 7598.
(9) Follis, A. V.; Galea, C. A.; Kriwacki, R. W. *Adv. Exp. Med. Biol.* **2012**, *725*, 27.
(10) Bonini, N. M.; Giasson, B. I. *Cell* **2005**, *123*, 359.
(11) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. *Annu. Rev. Biophys.* **2008**, *37*, 215.
(12) Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L. *Curr. Opin. Struct. Biol.* **2008**, *18*, 756.
(13) Kurzbach, D.; Platzer, G.; Schwarz, T. C.; Henen, M. A.; Konrat, R.; Hinderberger, D. *Biochemistry (Mosc.)* **2013**, *52*, 5167.
(14) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. *J. Mol. Graph. Model.* **2001**, *19*, 26.
(15) Uversky, V. N. *Protein Sci. Publ. Protein Soc.* **2002**, *11*, 739.
(16) Van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M. *Chem. Rev.* **2014**, *114*, 6589.
(17) Konrat, R. *J. Magn. Reson. San Diego Calif 1997* **2014**, *241*, 74.
(18) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* **2007**, *447*, 1021.
(19) Sibille, N.; Bernadó, P. *Biochem. Soc. Trans.* **2012**, *40*, 955.
(20) Schuler, B.; Hofmann, H. *Curr. Opin. Struct. Biol.* **2013**, *23*, 36.
(21) Rajendran, A.; Endo, M.; Sugiyama, H. *Adv. Protein Chem. Struct. Biol.* **2012**, *87*, 5.
(22) Felli, I. C.; Pierattelli, R. *IUBMB Life* **2012**, *64*, 473.
(23) Lipari, G.; Szabo, A. *Biochemistry (Mosc.)* **1981**, *20*, 6250.
(24) Peng, J. W.; Wagner, G. *J. Magn. Reson. 1969* **1992**, *98*, 308.
(25) Peng, J. W.; Wagner, G. *Biochemistry (Mosc.)* **1992**, *31*, 8571.
(26) Bertoncini, C. W.; Jung, Y.-S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 1430.
(27) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111.
(28) Boisbouvier, J.; Bax, A. *J. Am. Chem. Soc.* **2002**, *124*, 11038.
(29) Bernadó, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 17002.

(30) Bernadó, P.; Bertoncini, C. W.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2005**, *127*, 17968.
(31) Carlier, L.; Balayssac, S.; Cantrelle, F.-X.; Khemtémourian, L.; Chassaing, G.; Joliot, A.; Lequin, O. *Biophys. J.* **2013**, *105*, 667.
(32) Kissinger, C. R.; Liu, B. S.; Martin-Blanco, E.; Kornberg, T. B.; Pabo, C. O. *Cell* **1990**, *63*, 579.
(33) Mesleh, M. F.; Veglia, G.; DeSilva, T. M.; Marassi, F. M.; Opella, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 4206.
(34) Mesleh, M. F.; Opella, S. J. *J. Magn. Reson. San Diego Calif 1997* **2003**, *163*, 288.
(35) Dunker, A. K.; Obradovic, Z.; Romero, P.; Garner, E. C.; Brown, C. J. *Genome Inform. Workshop Genome Inform.* **2000**, *11*, 161.
(36) Brown, C. J.; Takayama, S.; Campen, A. M.; Vise, P.; Marshall, T. W.; Oldfield, C. J.; Williams, C. J.; Dunker, A. K. *J. Mol. Evol.* **2002**, *55*, 104.
(37) Chen, J. W.; Romero, P.; Uversky, V. N.; Dunker, A. K. *J. Proteome Res.* **2006**, *5*, 888.
(38) Tompa, P. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **2003**, *25*, 847.
(39) Diella, F.; Haslam, N.; Chica, C.; Budd, A.; Michael, S.; Brown, N. P.; Trave, G.; Gibson, T. J. *Front. Biosci. J. Virtual Libr.* **2008**, *13*, 6580.
(40) Schroeder, R.; Barta, A.; Semrad, K. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 908.
(41) Tompa, P.; Csermely, P. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **2004**, *18*, 1169.
(42) Rogers, J. M.; Steward, A.; Clarke, J. *J. Am. Chem. Soc.* **2013**, *135*, 1415.
(43) Chan, S. K.; Pöpperl, H.; Krumlauf, R.; Mann, R. S. *EMBO J.* **1996**, *15*, 2476.
(44) Cumberworth, A.; Lamour, G.; Babu, M. M.; Gsponer, J. *Biochem. J.* **2013**, *454*, 361.
(45) Daniels, A. J.; Williams, R. J.; Wright, P. E. *Neuroscience* **1978**, *3*, 573.
(46) Holt, C. *Curr. Opin. Struct. Biol.* **2013**, *23*, 420.
(47) Krumlauf, R. *Cell* **1994**, *78*, 191.
(48) Stern, C. D.; Foley, A. C. *Cell* **1998**, *94*, 143.
(49) Kornberg, T. B. *J. Biol. Chem.* **1993**, *268*, 26813.
(50) Spatazza, J.; Di Lullo, E.; Joliot, A.; Dupont, E.; Moya, K. L.; Prochiantz, A. *Pharmacol. Rev.* **2013**, *65*, 90.
(51) Passner, J. M.; Ryoo, H. D.; Shen, L.; Mann, R. S.; Aggarwal, A. K. *Nature* **1999**, *397*, 714.
(52) Piper, D. E.; Batchelor, A. H.; Chang, C. P.; Cleary, M. L.; Wolberger, C. *Cell* **1999**, *96*, 587.
(53) Qian, Y. Q.; Resendez-Perez, D.; Gehring, W. J.; Wüthrich, K. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 4091.
(54) Gehring, W. J. *Science* **1987**, *236*, 1245.
(55) Alexandre, C.; Vincent, J.-P. *Dev. Camb. Engl.* **2003**, *130*, 729.
(56) Augustyniak, R.; Balayssac, S.; Ferrage, F.; Bodenhausen, G.; Lequin, O. *Biomol. NMR Assign.* **2011**, *5*, 229.
(57) Marley, J.; Lu, M.; Bracken, C. *J. Biomol. NMR* **2001**, *20*, 71.
(58) Douzou, P. *Comptes Rendus Séances Société Biol. Ses Fil.* **1980**, *174*, 574.
(59) Jensen, M. R.; Ruigrok, R. W. H.; Blackledge, M. *Curr. Opin. Struct. Biol.* **2013**, *23*, 426.
(60) Palmer, A. G. *Chem. Rev.* **2004**, *104*, 3623.
(61) *Protein NMR spectroscopy: principles and practice*; Cavanagh, J., Ed.; 2nd ed.; Academic Press: Amsterdam ; Boston, 2007.
(62) Jarymowycz, V. A.; Stone, M. J. *Chem. Rev.* **2006**, *106*, 1624.

(63) Pelupessy, P.; Espallargas, G. M.; Bodenhausen, G. *J. Magn. Reson. San Diego Calif 1997* **2003**, *161*, 258.
(64) Pelupessy, P.; Ferrage, F.; Bodenhausen, G. *J. Chem. Phys.* **2007**, *126*, 134508.
(65) Abragam, A. *The principles of nuclear magnetism*; Clarendon Press ; Oxford University Press: Oxford [Oxfordshire]; New York, 1983.
(66) Farrow, N. A.; Zhang, O.; Szabo, A.; Torchia, D. A.; Kay, L. E. *J. Biomol. NMR* **1995**, *6*, 153.
(67) Kroenke, C. D.; Loria, J. P.; Lee, L. K.; Rance, M.; Palmer, A. G. *J. Am. Chem. Soc.* **1998**, *120*, 7905.
(68) Tugarinov, V.; Liang, Z.; Shapiro, Y. E.; Freed, J. H.; Meirovitch, E. *J. Am. Chem. Soc.* **2001**, *123*, 3055.
(69) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546.
(70) Barbato, G.; Ikura, M.; Kay, L. E.; Pastor, R. W.; Bax, A. *Biochemistry (Mosc.)* **1992**, *31*, 5269.
(71) Modig, K.; Poulsen, F. M. *J. Biomol. NMR* **2008**, *42*, 163.
(72) Houben, K.; Blanchard, L.; Blackledge, M.; Marion, D. *Biophys. J.* **2007**, *93*, 2830.
(73) Buevich, A. V.; Baum, J. *J. Am. Chem. Soc.* **1999**, *121*, 8671.
(74) Lemaster, D. M. *J. Biomol. NMR* **1995**, *6*, 366.
(75) .
(76) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197.
(77) Kadeřávek, P.; Zapletal, V.; Rabatinová, A.; Krásný, L.; Sklenář, V.; Žídek, L. *J. Biomol. NMR* **2014**, *58*, 193.
(78) Ishima, R.; Nagayama, K. *Biochemistry (Mosc.)* **1995**, *34*, 3162.
(79) Hill, R. B.; Bracken, C.; DeGrado, W. F.; Palmer, A. G. *J. Am. Chem. Soc.* **2000**, *122*, 11610.
(80) Ropars, V.; Bouguet-Bonnet, S.; Auguin, D.; Barthe, P.; Canet, D.; Roumestand, C. *J. Biomol. NMR* **2007**, *37*, 159.
(81) Ryabov, Y. E.; Fushman, D. *J. Am. Chem. Soc.* **2007**, *129*, 3315.
(82) Xue, Y.; Skrynnikov, N. R. *J. Am. Chem. Soc.* **2011**, *133*, 14614.
(83) Prompers, J. J.; Brüschweiler, R. *J. Am. Chem. Soc.* **2002**, *124*, 4522.
(84) Kay, L. E.; Nicholson, L. K.; Delaglio, F.; Bax, A.; Torchia, D. . *J. Magn. Reson. 1969* **1992**, *97*, 359.
(85) Ferrage, F.; Cowburn, D.; Ghose, R. *J. Am. Chem. Soc.* **2009**, *131*, 6048.
(86) Neri, D.; Billeter, M.; Wider, G.; Wüthrich, K. *Science* **1992**, *257*, 1559.
(87) Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13099.
(88) Schwalbe, H.; Fiebig, K. M.; Buck, M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A.; Glaser, S. J.; Smith, L. J.; Dobson, C. M. *Biochemistry (Mosc.)* **1997**, *36*, 8977.
(89) Hubbell, W. L.; Gross, A.; Langen, R.; Lietzow, M. A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 649.
(90) Berliner, L. J.; Grunwald, J.; Hankovszky, H. O.; Hideg, K. *Anal. Biochem.* **1982**, *119*, 450.
(91) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 158.
(92) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 170.
(93) Valérie Belle. Site-directed spin-labeling EPR spectroscopy.
(94) Nabuurs, S. M.; de Kort, B. J.; Westphal, A. H.; van Mierlo, C. P. M. *Eur. Biophys. J. EBJ* **2010**, *39*, 689.
(95) Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3.

(96) Huang, H.; Cafiso, D. S. *Biochemistry (Mosc.)* **2008**, *47*, 12380.
(97) Desjardins, G.; Meeker, C. A.; Bhachech, N.; Currie, S. L.; Okon, M.; Graves, B. J.; McIntosh, L. P. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 11019.
(98) Valeur, B. *Molecular fluorescence: principles and applications*; Wiley-VCH: Weinheim ; New York, 2002.
(99) Foucher, I.; Montesinos, M. L.; Volovitch, M.; Prochiantz, A.; Trembleau, A. *Dev. Camb. Engl.* **2003**, *130*, 1867.
(100) Van Dijk, M. A.; Murre, C. *Cell* **1994**, *78*, 617.
(101) Hill, A. V. *J. Physiol.* **1910**, *40*, 389.
(102) Hulme, E. C.; Trevethick, M. A. *Br. J. Pharmacol.* **2010**, *161*, 1219.
(103) Eaton, G. R. *Foundations of modern EPR*; World Scientific: Singapore ; River Edge, NJ, 1998.
(104) Belle, V.; Fournel, A.; Woudstra, M.; Ranaldi, S.; Prieri, F.; Thomé, V.; Currault, J.; Verger, R.; Guigliarelli, B.; Carrière, F. *Biochemistry (Mosc.)* **2007**, *46*, 2205.
(105) Altenbach, C.; Greenhalgh, D. A.; Khorana, H. G.; Hubbell, W. L. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 1667.
(106) Jeschke, G. *Chemphyschem Eur. J. Chem. Phys. Phys. Chem.* **2002**, *3*, 927.
(107) Mchaourab, H. S.; Lietzow, M. A.; Hideg, K.; Hubbell, W. L. *Biochemistry (Mosc.)* **1996**, *35*, 7692.
(108) Mileo, E.; Lorenzi, M.; Erales, J.; Lignon, S.; Puppo, C.; Le Breton, N.; Etienne, E.; Marque, S. R. A.; Guigliarelli, B.; Gontero, B.; Belle, V. *Mol. Biosyst.* **2013**, *9*, 2869.
(109) Borbat, P. P.; Costa-Filho, A. J.; Earle, K. A.; Moscicki, J. K.; Freed, J. H. *Science* **2001**, *291*, 266.
(110) Xue, Y.; Podkorytov, I. S.; Rao, D. K.; Benjamin, N.; Sun, H.; Skrynnikov, N. R. *Protein Sci.* **2009**, *18*, 1401.
(111) Tompa, P.; Fuxreiter, M. *Trends Biochem. Sci.* **2008**, *33*, 2.
(112) Mehring, M.; Waugh, J. *Phys. Rev. B* **1972**, *5*, 3459.
(113) Prestegard, J. H.; Bougault, C. M.; Kishore, A. I. *Chem. Rev.* **2004**, *104*, 3519.
(114) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 8407.
(115) Tjandra, N.; Omichinski, J. G.; Gronenborn, A. M.; Clore, G. M.; Bax, A. *Nat. Struct. Biol.* **1997**, *4*, 732.
(116) Mohana-Borges, R.; Goto, N. K.; Kroon, G. J. A.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2004**, *340*, 1131.
(117) Jensen, M. R.; Markwick, P. R. L.; Meier, S.; Griesinger, C.; Zweckstetter, M.; Grzesiek, S.; Bernadó, P.; Blackledge, M. *Struct. Lond. Engl. 1993* **2009**, *17*, 1169.
(118) Meier, S.; Grzesiek, S.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 9799.
(119) Obolensky, O. I.; Schlepckow, K.; Schwalbe, H.; Solov'yov, A. V. *J. Biomol. NMR* **2007**, *39*, 1.
(120) Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2008**, *130*, 11266.
(121) Bax, A. *Protein Sci. Publ. Protein Soc.* **2003**, *12*, 1.
(122) Wang L., W. H. Residual dipolar couplings: Measurements and applications to biomolecular studies., 2006.
(123) Fleming, K.; Matthews, S. *Methods Mol. Biol. Clifton NJ* **2004**, *278*, 79.
(124) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 9279.
(125) Hansen, M. R.; Mueller, L.; Pardi, A. *Nat. Struct. Biol.* **1998**, *5*, 1065.

(126) Ottiger, M.; Delaglio, F.; Bax, A. *J. Magn. Reson. San Diego Calif 1997* **1998**, *131*, 373.
(127) Permi, P.; Rosevear, P. R.; Annila, A. *J. Biomol. NMR* **2000**, *17*, 43.
(128) Yang, D.; Tolman, J. R.; Goto, N. K.; Kay, L. E. *J. Biomol. NMR* **1998**, *12*, 325.
(129) Chou, J. J.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **2000**, *18*, 101.
(130) Jaroniec, C. P.; Ulmer, T. S.; Bax, A. *J. Biomol. NMR* **2004**, *30*, 181.
(131) Cordier, F.; Dingley, A. J.; Grzesiek, S. *J. Biomol. NMR* **1999**, *13*, 175.

Appendix

Nanosecond Time Scale Motions in Proteins Revealed by High-Resolution NMR Relaxometry

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

Nanosecond Time Scale Motions in Proteins Revealed by High-Resolution NMR Relaxometry

Cyril Charlier,[†,⊥] Shahid Nawaz Khan,[†,⊥] Thorsten Marquardsen,[‡] Philippe Pelupessy,[†] Volker Reiss,[‡] Dimitris Sakellariou,[†,§] Geoffrey Bodenhausen,[†,∥] Frank Engelke,[‡] and Fabien Ferrage*[,†]

[†]Laboratoire des Biomolécules, Département de Chimie, UMR 7203 CNRS-UPMC-ENS, Ecole Normale Supérieure, 24 Rue Lhomond, 75231 Paris Cedex 05, France

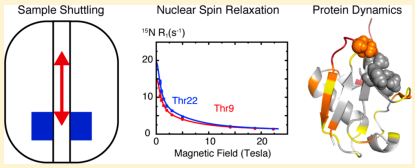
[‡]Bruker BioSpin GmbH, Silberstreifen 4, D 76287 Rheinstetten, Germany

[§]Laboratoire Structure et Dynamique par Résonance Magnétique, UMR 3299-SIS2M CEA/CNRS, IRAMIS, DSM, CEA Saclay, F-91191, Gif-sur-Yvette Cedex, France

[∥]Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne, BCH, 1015 Lausanne, Switzerland

Supporting Information

ABSTRACT: Understanding the molecular determinants underlying protein function requires the characterization of both structure and dynamics at atomic resolution. Nuclear relaxation rates allow a precise characterization of protein dynamics at the Larmor frequencies of spins. This usually limits the sampling of motions to a narrow range of frequencies corresponding to high magnetic fields. At lower fields one cannot achieve sufficient sensitivity and resolution in NMR. Here, we use a fast shuttle device where the polarization builds up and the signals are detected at high field, while longitudinal relaxation takes place at low fields $0.5 < B_0 < 14.1$ T. The sample is propelled over a distance up to 50 cm by a blowgun-like system in about 50 ms. The analysis of nitrogen-15 relaxation in the protein ubiquitin over such a wide range of magnetic fields offers unprecedented insights into molecular dynamics. Some key regions of the protein feature structural fluctuations on nanosecond time scales, which have so far been overlooked in high-field relaxation studies. Nanosecond motions in proteins may have been underestimated by traditional high-field approaches, and slower supra-τ_c motions that have no effect on relaxation may have been overestimated. High-resolution relaxometry thus opens the way to a quantitative characterization of nanosecond motions in proteins.



INTRODUCTION

The chemical and physical principles underlying protein function can only be unraveled by gaining insight into both structural and dynamic features. Nuclear magnetic resonance spectroscopy is unmatched in its ability to provide such insight at atomic resolution. Nuclear spin relaxation, i.e., the return of the perturbed magnetization to its equilibrium, allows one to characterize internal protein dynamics in two distinct ranges: fast pico- to nanosecond and slow micro- to millisecond time scales.[1] Overall rotational diffusion of proteins, which occurs typically on time scales τ_c on the order of a few nanoseconds to several tens of nanoseconds, and internal motions on time scales $\tau_{int} \lesssim \tau_c$ lead to stochastic fluctuations of orientation-dependent spin interactions such as dipole−dipole couplings and anisotropic chemical shifts. The resulting relaxation rates depend on spectral density functions $J(\omega)$, which are defined as Fourier transforms of the correlation functions $C(t)$ of the orientation-dependent interactions. The measurement of a series of relaxation rates allows one to "map" the spectral density functions $J(\omega)$,[2] thus providing quantitative parameters for models of overall and internal motions. Most models employed to date rely on the hypothesis that internal motions are statistically independent of the rotational diffusion of a macromolecule, which is usually justified by the separation of the time scales of overall and internal motions, the latter being hitherto considered to be at least an order of magnitude faster than the former.[3] Internal motions are usually described by one or two order parameters and discrete internal correlation times[4,5] or by a distribution of such internal correlation times.[6−8] The hypothesis that the time scales of overall and local motions can be separated and that they are statistically independent has however been challenged.[9] The discrimination between different models of motions has proven to be difficult because the spectral density functions could only be properly determined over few narrow ranges of fairly high frequencies. This limitation can be overcome by "relaxometry", i.e., by measuring relaxation rates over a wide range of magnetic fields, typically from 1 μT to 3 T. So far, this could only be achieved at the expense of sensitivity and resolution.[10,11] Relaxometry can

Received: September 26, 2013
Published: November 14, 2013

dx.doi.org/10.1021/ja409820g | J. Am. Chem. Soc. 2013, 135, 18665−18672

be reconciled with high-resolution high-field NMR by rapidly "shuttling" the sample from high to low field and back.[12−14] Slower shuttling has allowed fruitful studies of slowly relaxing phosphorus-31 and carbon-13 nuclei in lipids,[15] but relaxation of nitrogen-15 in proteins is so fast at low fields that the shuttling must be carried out very rapidly. To the best of our knowledge, there is only one pioneering study,[16] albeit shuttling was limited to a fairly narrow range of magnetic fields (down to 4 T).

Here, we use high-resolution relaxometry combined with traditional high-field measurements to measure relaxation rates over nearly 2 orders of magnitude (0.5−22.3 T.) We illustrate the power of this method by revealing internal nanosecond-time scale dynamics in the protein ubiquitin. (Poly)-ubiquitination is a mechanism involved in many biological cell-signaling processes, from protein degradation to DNA damage response. Signaling is mediated by interactions of ubiquitin and polyubiquitin with a broad range of protein partners. Such a diversity is made possible through the conformational flexibility of its 76 amino acid chain.[17] This observation was highlighted in a recent study[18] where the internal dynamics of ubiquitin were modified to bind selectively to a single partner, primarily by reducing the flexibility of its β_1−β_2 turn (residues 7−13).

It has been shown that the intrinsic flexibility of ubiquitin in its free apo form leads to a rich conformational landscape, which is similar to the conformational diversity in ubiquitin complexes.[19] Binding to a given partner can be described by an induced fit, a conformational selection, or an intermediate mechanism, depending on the time scales of conformational transitions and the lifetimes of encounter complexes.[20] Unraveling the time scales of internal motions in ubiquitin is obviously a prerequisite to understanding the kinetic pathways of binding reactions. Many NMR studies have sought to identify signatures of chemical exchange in ubiquitin in solution,[21−25] demonstrating the presence of internal motions on slow time scales ($10 \lesssim \tau_{int} \lesssim 100$ μs). Residual dipolar couplings (RDCs)[26−30] also point to the presence of extensive slow supra-τ_c time scales of 5 ns $\lesssim \tau_{int} \lesssim$ 10 ms (with $\tau_c \approx 5$ ns), which could not be detected by high-field relaxation or by chemical exchange phenomena. Several studies of faster sub-τ_c motions with $\tau_{int} \lesssim \tau_c = 5$ ns have been carried out using ^{15}N or ^{13}C relaxation.[31−34] Most of these studies rely on data acquired at a single high magnetic field and use simple spectral density functions. The results are often compared to those derived from RDC measurements, and discrepancies are attributed to contributions of supra-τ_c motions. We show that high-resolution relaxometry can reveal a surprising complexity of internal protein motions with time scales comparable to overall tumbling that were overlooked in high-field relaxation studies. We show that slower supra-τ_c motions have likely been overestimated so far, in particular in the essential extended β_1−β_2 turn.

RESULTS AND DISCUSSION

The sensitivity of relaxometry to motions on nanosecond time scales is illustrated in Figure 1. We have simulated the longitudinal relaxation rates of nitrogen-15 at several magnetic fields, assuming motions over a range of time scales and amplitudes. An extended model-free[5] spectral density function was assumed with variable order parameters and time scales for slow motions on nanosecond time scales. Analytical expressions can be found in the Supporting Information. The longitudinal

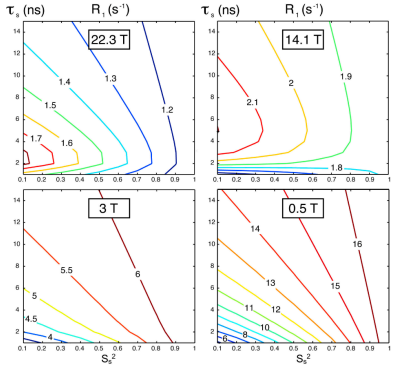


Figure 1. Simulated dependence of longitudinal nitrogen-15 relaxation rates on internal motions with nanosecond time scales, i.e., below and above the correlation time for overall tumbling $\tau_c = 5$ ns. An extended model-free spectral density function was used with the following parameters: $\tau_c = 5$ ns; correlation time for fast internal motions $\tau_{fast} = 10$ ps; order parameters for fast internal motions $S^2_{fast} = 0.8$; order parameters for slow internal motions $0.1 < S^2_{slow} < 1.0$ (x axis); 1 ns < τ_{int} < 15 ns (y axis).

relaxation rates vary strongly with the magnetic field, depending on the parameters of local motions. Both high- and low-field relaxation rates are remarkably sensitive to slow sub-τ_c motions if the order parameters are low. The rates at low fields are more sensitive to internal motions when their correlation times are comparable to the overall tumbling time. Relaxation rates at 0.5 T are more sensitive to internal motions than those recorded at 3 T, which underscores the advantages of studying relaxation at low magnetic fields. Longitudinal relaxation rates are largely insensitive to internal motions with slow time scales $\tau_{int} \gtrsim \tau_c/2$ at 14.1 T, i.e., at high magnetic fields where many studies of internal dynamics in ubiquitin have been carried out so far. The different patterns for the dependence of longitudinal relaxation rates at high and low fields underline the enhanced sensitivity to nanosecond motions of relaxation measurements over a broad range of magnetic fields.

We developed a pneumatic system for fast shuttling, based on a system that was originally developed for liquid-state dynamic nuclear polarization (DNP) studies where the proton polarization observed at 14.1 T can be enhanced by saturating EPR transitions at 0.34 T.[35] Our shuttle consists of a custom-designed probe (Figure 2a,b), a transfer system, and a control unit, as described in more detail in Supporting Information. The probe uses two saddle coils, like in standard high-resolution probes. The inner coil is doubly tuned for ^{13}C and ^{15}N, while the orthogonal outer coil is doubly tuned for ^{1}H and ^{2}H. This reduces interactions between the sample and the electric component of the rf field, albeit at the expense of a slight loss of sensitivity. The design affords a spectral resolution and line shapes comparable to those obtained using state-of-the-art high-resolution probes at 600 MHz. Special care was taken (Figure 2a) to reduce vibrations arising from the abrupt "landing" of the shuttle at the lower end (see Supporting Information). A long tube guides the shuttle during its motion

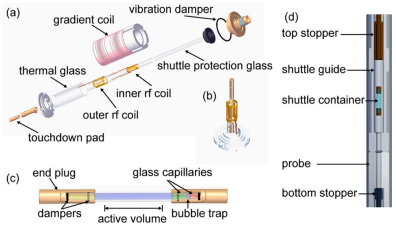


Figure 2. Schematic views of the fast shuttling system: (a) expanded view of the upper part of the probe; (b) coil assembly with shuttle container; (c) shuttle container; (d) schematic view (not to scale) of the full assembly in the bore of a high-field magnet.

(Figure 2c). The upper position of the shuttle is controlled by an adjustable inner tube. The inner tube is connected to the guiding tube by another damping system to reduce vibrations.

A special quartz container was chosen for protein samples (Figure 2d). Synthetic amorphous quartz glass with a low magnetic susceptibility can resist a large number of shocks. (A single container was used for more than 500 000 shuttling events in the course of this study.) In the shuttle container, a ~100 μL sample compartment is separated from a ~10 μL "bubble catcher" compartment by a narrow capillary (150 μm inner diameter). Bubbles appearing in the course of extensive experimental series can be centrifuged into the bubble catcher before resuming the experiments. The active volume that lies within the rf coils is 60 μL (Figure 2d). The resulting sensitivity of this system is about an order of magnitude lower than the sensitivity of a room-temperature TXI probe used with a large-volume sample. This currently limits applications to the study of (bio)molecules that have a good solubility or favorable relaxation properties.

We used this shuttling system to measure the longitudinal relaxation of backbone nitrogen-15 nuclei in uniformly nitrogen-15 labeled human ubiquitin in acetate buffer at pH 4.5 and T = 296.6 ± 0.6 K. The pulse sequence used for these measurements is shown in Figure 3. After a recovery delay in high field (B_0^{high} = 14.1 T) to allow the Boltzmann polarization to build up, the temperature to be regulated, and the field-frequency lock to stabilize the field, the longitudinal polarization N_z of nitrogen-15 is enhanced using the refocused INEPT method.[36,37] The sample is then transferred in 41 < τ_{up} < 54 ms to a predetermined position in the stray field 27 < z < 46 cm above the magnetic center. The polarization is then allowed to relax in a low field B_0^{low} for a duration T_{rel} and transferred back to high field B_0^{high} in 40 < τ_{down} < 70 ms. During a stabilization delay τ_{st} = 100 ms, convection and vibrations are allowed to settle. Finally, the longitudinal nitrogen-15 polarization N_z is converted back into transverse proton magnetization for detection. The average signal-to-noise ratios in two-dimensional (2D) spectra obtained for the shortest relaxation delays T_{rel} = 39 ms were S/N = 66 at B_0^{low} = 5 T (at z = 27 cm above the magnetic center) and S/N = 24 for T_{rel} = 51 ms at B_0^{low} = 0.5 T (z = 46 cm) when 16 transients were recorded for each of 64 complex points in the indirect t_1 dimension (the experimental time was 85 min for each interval T_{rel}).

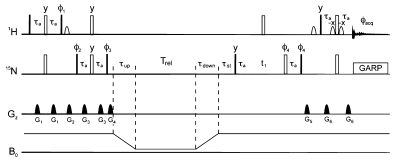


Figure 3. Pulse sequence for the measurement of longitudinal nitrogen-15 relaxation of amide nitrogen nuclei in proteins at various low fields B_0^{low} and recovery and detection at high field B_0^{high}. Narrow (filled) and wide (open) rectangles represent 90° and 180° pulses, respectively. Pulse phases are along the x-axis of the rotating frame unless otherwise specified. The bell-shaped pulses represent 1.2 ms sinc pulses. All delays τ_a are set to $1/|4J_{NH}|$, with J_{NH} = −92 Hz. The stabilization delay τ_{st} = 100 ms allows for convection currents and vibrations to settle. Pulsed field gradients G_z have smoothed rectangular amplitude profiles and 1 ms durations. Their peak amplitudes are G_1 = 25, G_2 = 40, G_3 = 11.5, G_4 = 20.5, G_5 = 40, G_6 = 15 G cm^{-1}. The phase cycles were φ_1 = {y, y, y, y, $-y$, $-y$, $-y$, $-y$}; φ_2 = {x, $-x$}; φ_3 = {y, y, y, y, y, y, y, $-y$, $-y$, $-y$, $-y$, $-y$, $-y$, $-y$, $-y$}; φ_4 = {x, x, $-x$, $-x$}; φ_{acq} = {x, $-x$, $-x$, x, $-x$, x, x, $-x$, $-x$, x, x, $-x$, x, $-x$, $-x$, x}.

Longitudinal relaxation rates were recorded at 7 different low magnetic fields. Each measurement was repeated 2 or 3 times. In addition, a full set of conventional nitrogen-15 relaxation experiments was recorded without shuttling at 14.1, 18.8, and 22.3 T (600, 800, and 950 MHz for ^{1}H) using state-of-the-art methods to cancel cross-correlation effects.[38−41] The parameters of rotational diffusion were derived from relaxation rates using the program ROTDIF.[42] Diffusion tensors obtained at all three high fields are virtually identical (see Supporting Information). On the basis of a solution-state structure of ubiquitin (PDB code 1d3z),[43] we obtained an axially symmetric diffusion tensor with $D_\parallel/D_\perp$ = 1.18 ± 0.08 and an overall correlation time $\tau_c = (6\mathrm{Tr}(\mathbf{D}))^{-1}$ = 4.84 ± 0.2 ns, where Tr($\mathbf{D}$) is the trace of the diffusion tensor from our measurements at 18.8 T. At a lower concentration (200 μM) at 14.1 T we found a slightly shorter overall correlation time τ_c = 4.22 ± 0.15 ns. The small variation of τ_c and the fact that the anisotropy and orientation of the diffusion tensor were found to be independent of concentration (see Supporting Information) suggest that the protein is in monomeric form at pH 4.5 and not in a monomer/dimer equilibrium, as found at neutral pH.[44] In our case, the low pH leads to the protonation of His68 and is likely to reduce the binding affinity of symmetric dimers that associate at the hydrophobic patch at neutral pH, as shown in diubiquitin.[45] Therefore, we believe that the slight increase of the overall correlation time τ_c was due to nonspecific intermolecular interactions at high concentration so that we could describe the overall rotational diffusion by a single time-independent tensor.

Particular care has been taken to control the temperature in these experiments, since temperature regulation at low field B_0^{low} is not yet feasible in our prototype. We have measured differences of chemical shifts between subsets of signals and calibrated them as a function of temperature.[23] The actual sample temperature was then derived from these chemical shift differences in each experiment. Only experiments with 296 ≤ T ≤ 297.2 K were retained in the analysis so that systematic errors can be safely neglected (see Supporting Information). Conventional high field-relaxation measurements at 14.1 and

18.8 T were carried out at 296.5 K. Rates measured at 22.3 T and 298.5 K were corrected to account for the change in τ_c.

The longitudinal relaxation rates $R_1(B_0) = 1/T_1(B_0)$ of ^{15}N nuclei in the backbone of ubiquitin determined at 10 different fields are shown in Figure 4. Cross-relaxation pathways lead to multiexponential decays, which in high fields can usually be transformed into monoexponential decays by suitable pulse sequences.[46,47] Since in our prototype it is not possible to apply any rf pulses in low fields, systematic deviations from monoexponential decays must be taken into account in the analysis.[16] We have developed a protocol dubbed "iterative correction for the analysis of relaxation under shuttling" (ICARUS). First, the analysis of relaxation rates was carried out (in terms of overall tumbling and microdynamics) at all three high fields 14.1, 18.8, and 22.3 T using the programs ROTDIF[42] and DYNAMICS.[48] The parameters resulting from this initial step were then used to predict the deviations from simple exponential decays in a spin system that comprises one ^{15}N−^{1}H pair and two remote protons (see Supporting Information). For each field B_0^{low}, we simulated the relaxation of each ^{15}N−^{1}H pair in ubiquitin *during* shuttling, with constant velocity (see Supporting Information). The deviations between the calculated "apparent" nitrogen-15 relaxation rates and the "true" low-field relaxation rates were then used to correct for systematic errors in the experimental rates. In a second iteration, longitudinal relaxation rates at all 10 fields and ^{15}N{^{1}H} NOEs at the three high fields 14.1, 18.8, and 22.3 T were used as input to the program DYNAMICS. This cycle was reiterated four times to achieve a satisfactory convergence for all residues. The typical corrections varied from 4.5% to 13%. Cross-correlations of the fluctuations of nitrogen-15 chemical shift anisotropies and ^{15}N−^{1}H dipolar couplings are dominant above 3 T, with average corrections ranging from 5.1% at 3 T to 9.2% at 5 T, while ^{15}N−^{1}H dipolar cross-relaxation dominates below 2 T, with corrections on the order of 11% at fields below 1 T.

The results of this analysis are shown in Figure 5. As expected from studies of ^{15}N and ^{13}C relaxation,[31,32] residual dipolar couplings,[26,28,29] and molecular dynamics,[49−51] we find that ubiquitin is fairly rigid. However, we determined the order parameters to be significantly lower than in earlier relaxation-based studies. We may compare (Figure 4) (i) the order parameters resulting from all 10 fields, (ii) those obtained from relaxation rates at 14.1 T only, and (iii) those obtained at three high fields (14.1, 18.8, and 22.3 T). With few exceptions, the order parameters resulting from our analysis of relaxation at 10 magnetic fields are the lowest. In particular, the dynamics of the crucial β_1−β_2 turn (residues 7−12) can be best described by an extended model-free[5] spectral density function with similar time scales for all six residues (see Supporting Information). A global fit of these six residues gives a common effective time scale τ_{7-12} = 2 ns. This is in good agreement with the well-documented hypothesis of a collective motion.[52] This motion was so far believed to occur on a much slower, so-called supra-τ_c time scale $\tau_{7-12} > \tau_c$, since RDC studies indicated large-amplitude motions while relaxation at a single high field failed to detect such motions. Interestingly, lower order parameters and motions on similar time scales (see Supporting Information) are also found at the C-terminus of helix α_1 and loop α_1−β_4 (residues 33 and 36), which participate, along with the β_1−β_2 turn, in the principal mode of ubiquitin dynamics.[19,53] The relaxation of this principal mode could be

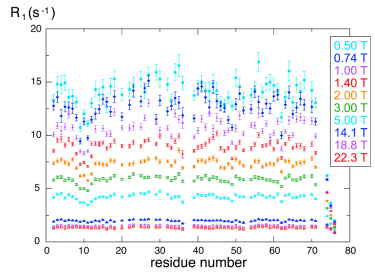


Figure 4. Experimental longitudinal relaxation rates $R_1 = 1/T_1$ of ^{15}N in backbone amide groups of ubiquitin as a function of magnetic fields $0.5 \le B_0^{low} \le 5$ T and $14.1 \le B_0^{high} \le 22.3$ T. From bottom to top: B_0^{high} = 22.3, 18.8, 14.1 (red squares, magenta crosses, blue triangles); B_0^{low} = 22.3, 18.8, 14.1 5.0, 3.0, 2.0, 1.4, 1.0, 0.74, and 0.5 T (cyan diamonds, green circles, orange squares, red crosses, magenta triangles, blue diamonds, cyan circles). Note that all rates increase with decreasing field. The lower B_0^{low} is, the greater are the variations of the rates along the backbone.

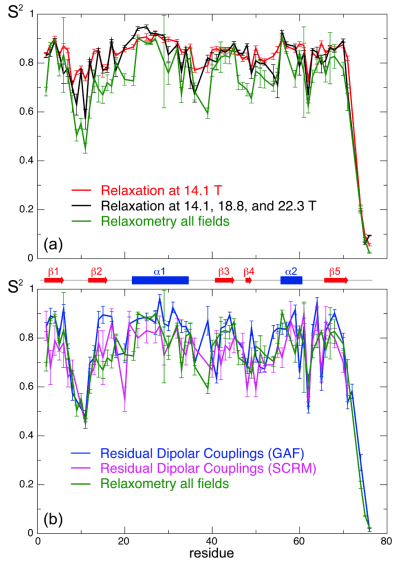


Figure 5. (a) Order parameters S^2 in ubiquitin obtained from the analysis of nitrogen-15 relaxation rates, taking into account (red) relaxation rates at 14.1 T only; (black) relaxation data at three fields 14.1, 18.8, and 22.3 T; (green) relaxation rates at all 10 fields from 0.5 to 22.3 T. (b) Comparison of order parameters obtained from relaxation rates at all fields (green) and from analysis of residual dipolar couplings (RDCs) in large sets of orienting media, either by GAF (blue)[30] or SCRM (purple).[28]

described in silico with two time scales, 0.4 and 13 ns.[53] Our analysis assumes only one single time scale for the semilocal motions, and the fit leads to an intermediate value, $\tau_{7-12} = 2$ ns, which could well result from effective averaging between these two time scales. More complex models of spectral density functions should open the way to a better agreement between experimental rates and theory.

The inability of high-field relaxation studies to identify these motions is hardly surprising, since most studies have only been carried out at a single field. Therefore, the sampling of the spectral density function was insufficient and did not characterize sufficiently well motions on nanosecond time scales. The consequences of undersampling of the spectral density function are exacerbated if one uses simple models that do not properly reproduce the actual spectral density functions. Strikingly, the order parameters for fast motions obtained by extended-model free analysis of all relaxation rates match very well with order parameters obtained from the analysis of relaxation data at 14.1 T only (see Supporting Information). This point is also nicely illustrated by the analysis of a 1.2 μs molecular dynamics trajectory of ubiquitin.[50] In this study, order parameters S^2 derived from the average orientations of NH vectors were low for the β_1–β_2 turn. However, when the nonexponential correlation functions were forced to fit with a simple extended model-free correlation function, the order parameters became significantly higher, similar to those found in relaxation studies at 14.1 T.[32]

Figure 5b presents the comparison of orders parameters obtained (i) from our relaxometry analysis of relaxation at 10 magnetic fields and (ii) from two independent analyses of residual dipolar couplings (RDCs) in large sets of oriented media using the Gaussian axial fluctuations (GAF)[30] or the self-consistent RDC-based model-free analysis (SCRM)[28] approaches. The three profiles are similar. In particular, the order parameters in the β_1–β_2 turn (residues 7–12) are almost identical so that one expects the amplitudes of motions in this turn that cannot be detected by relaxation to be very small. Significant correlations between slow supra-τ_c motions of the β_1–β_2 turn and those of the β sheet are therefore unlikely, although correlated motions in the core of the β-sheet cannot be excluded.[52] Our relaxometry data show that the whole β_2 strand, which lies at the edge of the β-sheet, is significantly dynamic. Similarly, studies of the third immunoglobin binding domain of streptococcal protein G (GB3) have also shown the presence of enhanced motions in the last strand of an otherwise fairly rigid β-sheet.[54] These results differ from those of a GAF analysis of dynamics in ubiquitin,[29] where the β_2-strand is found to be rigid. However, they are in better agreement with results from the SCRM analysis.[28]

In addition to this region, our relaxometry method allowed us to detect enhanced dynamics in several loops: β_2–α_1, β_4–α_2, and α_2–β_5 as well as in the β_3–β_4 turn and β_4 strand, which lies at the opposite edge of the β sheet. This is, again, in good qualitative agreement with both RDC studies (see Figure 5b). The agreement with accelerated molecular dynamics simulations is excellent, with a good correlation coefficient (R = 0.91) between the two data sets (see Supporting Information). Note, however, that the order parameters found by relaxometry are systematically (albeit only slightly) lower than those obtained by molecular dynamics.

Figure 6 shows a graphical representation of order parameters in ubiquitin. The edges of the β-sheet are occupied by the β_1–β_2 turn and the flexible C-terminal tail at one end and by the α_2–β_5 loop and β_3–β_4 turn at the other end, with the β_2 and β_4 strands on each side. All of these regions are found to be dynamic, albeit to different extent. The picture that emerges is a hierarchy of time scales[55] near the main binding interface of ubiquitin that consists of a β-sheet with a core that is flexible on a slow time scale of about 50 μs,[23,25] while its edges are mobile on faster nanosecond time scales. Between these two time scales, small correlated fluctuations of the β-sheet also appear to be allowed.[52] The ability of the edges of the interface to undergo conformational rearrangements on nanosecond time scales would be compatible with an induced fit mechanism in the early stage of binding.

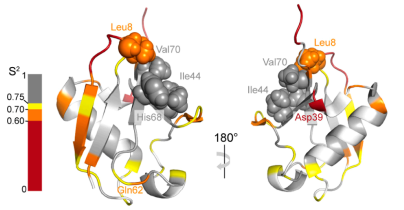


Figure 6. Large amplitude dynamics in ubiquitin on fast nanosecond time scales. Rigid residues with high order parameters $S^2 > 0.75$ are shown in gray. Mobile residues with intermediate and small order parameters are shown in yellow ($0.70 < S^2 < 0.75$), orange ($0.60 < S^2 < 0.70$), and red ($S^2 < 0.6$.). Residues for which no data are available are shown in white. The main interface with binding partners comprises the side chains of residues Leu8, Ile44, His68, and Val70 represented by space-filling models.

Figure 7 shows a few selected plots of longitudinal relaxation rates $R_1(B_0)$. Dramatic differences can be observed between mobile and rigid residues. A good agreement between experimental and theoretical profiles is observed for most residues. Some dispersion profiles feature systematic discrepancies, which highlight the limitations of current models of

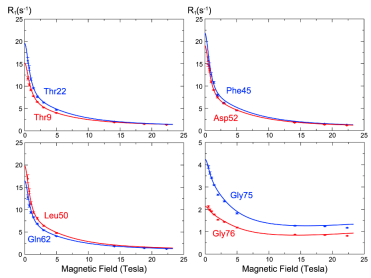


Figure 7. Longitudinal relaxation rates $R_1(B_0)$ as a function of the static field (so-called "relaxometry dispersion profiles") for eight selected residues in ubiquitin. Note that the vertical scale is expanded by a factor of 5 for the two C-terminal glycines (bottom right). The blue and red dots show corrected longitudinal relaxation rates, adjusted to compensate for relaxation during shuttling, while the lines show dispersion profiles calculated from the microdynamic parameters obtained in our analysis.

spectral density functions. Our experimental data call for the development of new, more sophisticated models. Some relaxation profiles, e.g., for the highly mobile C-terminal glycine residues G75 and G76, present deviations from the theoretical profiles at both high and low magnetic fields, even when postulating a spectral density function comprising a sum of three Lorentzian functions with five adjustable parameters, thus suggesting the presence of a distribution of time scales.[8] In particular, with only two time scales for internal motions, the fitted spectral density function is rather flat at high frequencies. Hence, contributions of the chemical shift anisotropy to relaxation lead to an increase in the $R_1(B_0)$ curve between 14 and 23 T, in contradiction with experimental results. Similarly, a small but systematic underestimation of the spectral density $J(\omega=0)$ in some of the most mobile regions, and hence of the back-predicted transverse relaxation rates (see Supporting Information), can be understood by postulating a rapid initial decay of the spectral density function at low frequencies. Interestingly, no significant contribution of chemical exchange to transverse relaxation R_{ex} could be detected (see Supporting Information).[39,56] The analysis of relaxation at three high fields also leads to underestimate $J(\omega=0)$, as illustrated by the need for urealistic R_{ex} contributions to fit *all* relaxation data. Nanosecond fluctuations of the overall diffusion tensor associated with motions of the C-terminal tail, transient oligomerization,[10] and mode coupling of local and global motions[9,57,58] may be responsible for these unexpected features.

We have measured and analyzed residue-specific relaxation rates in a protein over a range of nearly 2 orders of magnitude of magnetic fields. Our high-resolution relaxometry approach reveals unexpected motions in the protein ubiquitin. In particular, the motion of the $\beta_1-\beta_2$ turn appears to have larger amplitudes than could be previously identified by relaxation at high fields, in agreement with RDCs and MD. Until now, discrepancies between high-field relaxation and RDC-based methods were attributed to the cutoff of internal motions by overall rotation. High-field relaxation studies have led to underestimate sub-τ_c and near-τ_c motions because relaxation rates in high fields are not sufficiently sensitive to motions in the nanosecond frequency range. Although many proteins, and ubiquitin in particular, are mobile on slow supra-τ_c time scales (slower than overall rotational diffusion), a mere comparison of order parameters obtained from high-field relaxation and RDCs is likely to overestimate such slow motions. This study shows that high-resolution relaxometry with fast sample shuttling allows one to map the spectral density functions in exquisite detail and offers unprecedented information about local motions in proteins on time scales that are faster than or comparable to their overall tumbling.

METHODS

Magnetic Field Mapping. The magnetic field was measured as a function of the height above the magnetic center in steps of 1 mm using a homemade mapping device with two calibrated triple-axes Hall probes (Senis) with a precision of 0.1%. A CH3A10mE3D transducer was used for measurements from 0.05 to 2 T, while a 03A05F-A20T0K5Q transducer was used between 1 and 13 T.

Relaxation Experiments. The experiments were carried out on samples of 0.2 and 3 mM uniformly ^{15}N labeled human ubiquitin (Giotto) in 50 mM acetate buffer (pH 4.5) in H_2O/D_2O (90/10 v/v) at $296 \leq T \leq 297.2$ K using a 600 MHz Bruker Avance III spectrometer equipped with our pneumatic sample shuttle for measurements at low field. The pulse sequence shown in Figure 3 was used for $0.5 < B_0 < 5$ T, with a recovery delay of 2.2 s. All experiments were acquired with 16 transients and 64 complex points in the indirect t_1 dimension. Water-flip back pulses were applied to minimize the saturation of the water resonance;[59] the WATERGATE[60] scheme was used prior to detection. Frequency sign discrimination in the ω_1 domain was achieved with the States−TPPI method.[61] A full relaxation decay comprising 7−8 interleaved spectra could be recorded in 10−12 h. All signals were recorded at 14.1 T with a prototype probe equipped with z axis gradients and processed and analyzed with NMRPipe.[62] The relaxation curves at low fields were fitted to monoexponential functions.

Relaxation Data Analysis. The analysis of relaxation rates was performed with our ICARUS process, which uses ROTDIF[42] and DYNAMICS[48] at each iteration. All programs are written in Matlab (MathWorks, Inc.) A full description is given in Supporting Information. In order to account for systematic errors, a jack-knife procedure was used: the analysis was repeated seven times while excluding one of the seven low-field relaxation rates. The order parameters shown in Figure 5 result from the average over the seven analyses, and the errors correspond to the standard deviation of all seven values weighted by $6^{1/2}$.

The analysis was carried out with similar parameters as in many other NMR studies of protein dynamics. An internuclear nitrogen−hydrogen distance d_{NH} = 1.02 Å and a ^{15}N chemical shift anisotropy of −160 ppm were assumed to be common to all peptide bonds.

The effective distances between the H^N amide proton and the two additional protons are critical for scaling the corrections in the ICARUS procedure. In order to determine these distances, we measured longitudinal relaxation rates at 14.1 T with experiments similar to the shuttling method but where the longitudinal nitrogen-15 polarization is allowed to evolve during fixed intervals before and after the relaxation delay, during which no rf pulses are applied. The optimal distances d_{HH} were found to vary between 1.6 and 2.7 Å with an average of 2.1 Å. This result is confirmed by a computation of the sum of dipolar interactions with all protons (see Supporting Information) where the median value corresponds to an effective distance d_{HH} = 2.07 Å. Unfortunately, site-specific variations of d_{HH} obtained in the two approaches were only weakly correlated so that we decided to use an average d_{HH} = 2.1 Å for all residues. In order to evaluate the potential systematic errors of the resulting order parameters, we carried out a complete ICARUS analysis for a series of distances $1.7 < d_{HH} < 2.6$ Å. Results are shown in the Supporting Information. Order parameters of some sites tend to be sensitive to the distance d_{HH}, but the main features of our analysis, in particular the low order parameters found in the $\beta_1-\beta_2$ turn, remain stable regardless of the distance d_{HH}.

ASSOCIATED CONTENT

Supporting Information

Detailed presentation of the shuttle system; temperature control; description of the ICARUS protocol; results of ICARUS analysis; tables of all measured and corrected relaxation rates as well as microdynamic parameters obtained from analysis of relaxation at 14.1 T only, three high fields, and all 10 fields. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

Fabien.Ferrage@ens.fr

Author Contributions

⊥C.C. and S.N.K. contributed equally.

Notes

The authors declare the following competing financial interest(s): T.M., V.R., and F.E. are employees of Bruker Biospin. The authors declare no other competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank A. Guiga (CEA) for machining the field-mapping device; Martin Blackledge (IBS, Grenoble), Arthur G. Palmer (Columbia University), Daniel Abergel (ENS), and Pau Bernado (CBS, Montpellier) for fruitful exchanges; Martin Blackledge for sharing data; and Daniel Abergel and Martin Blackledge for carefully reading the manuscript. We thank Philip Lottmann and Christian Griesinger (Max Planck Institute, Göttingen) for their contributions to the shuttle design as well as Nelly Morellet (ICSN, Gif-sur-Yvette) for assistance with high-field spectrometers. This research has received funding from the European Research Council (ERC) under the European Community's Seventh Framework Programme (FP7/2007-2013), ERC Grant 205119 (REvolutioNMR) to D.S. and Grant 279519 (2F4BIODYN) to F.F., as well as from the Agence Nationale de la Recherche (ANR-11-BS07-031-01). Financial support from the TGIR-RMN-THC Fr3050 CNRS is gratefully acknowledged.

■ REFERENCES

(1) Palmer, A. G. *Chem. Rev.* **2004**, *104*, 3623.
(2) Peng, J. W.; Wagner, G. *J. Magn. Reson.* **1992**, *98*, 308.
(3) Prompers, J. J.; Bruschweiler, R. *J. Am. Chem. Soc.* **2002**, *124*, 4522.
(4) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546.
(5) Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1990**, *112*, 4989.
(6) Calligari, P.; Calandrini, V.; Kneller, G. R.; Abergel, D. *J. Phys. Chem. B* **2011**, *115*, 12370.
(7) Ochsenbein, F.; Neumann, J. M.; Guittet, E.; Van Heijenoort, C. *Protein Sci.* **2002**, *11*, 957.
(8) Buevich, A. V.; Baum, J. *J. Am. Chem. Soc.* **1999**, *121*, 8671.
(9) Tugarinov, V.; Liang, Z. C.; Shapiro, Y. E.; Freed, J. H.; Meirovitch, E. *J. Am. Chem. Soc.* **2001**, *123*, 3055.
(10) Luchinat, C.; Parigi, G. *J. Am. Chem. Soc.* **2007**, *129*, 1055.
(11) Persson, E.; Halle, B. *J. Am. Chem. Soc.* **2008**, *130*, 1774.
(12) (a) Redfield, A. G. *Magn. Reson. Chem.* **2003**, *41*, 753. (b) Redfield, A. G. *J. Biomol. NMR* **2012**, *52*, 159.
(13) Victor, K.; Kavolius, V.; Bryant, R. G. *J. Magn. Reson.* **2004**, *171*, 253.
(14) Chou, C. Y.; Chu, M. L.; Chang, C. F.; Huang, T. H. *J. Magn. Reson.* **2012**, *214*, 302.
(15) Roberts, M. F.; Redfield, A. G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17066.
(16) Clarkson, M. W.; Lei, M.; Eisenmesser, E. Z.; Labeikovsky, W.; Redfield, A.; Kern, D. *J. Biomol. NMR* **2009**, *45*, 217.
(17) Dikic, I.; Wakatsuki, S.; Walters, K. J. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 659.
(18) Zhang, Y. N.; Zhou, L. J.; Rouge, L.; Phillips, A. H.; Lam, C.; Liu, P.; Sandoval, W.; Helgason, E.; Murray, J. M.; Wertz, I. E.; Corn, J. E. *Nat. Chem. Biol.* **2013**, *9*, 51.
(19) Lange, O. F.; Lakomek, N. A.; Fares, C.; Schroder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmuller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471.
(20) Miloushev, V. Z.; Levine, J. A.; Arbing, M. A.; Hunt, J. F.; Pitt, G. S.; Palmer, A. G. *J. Biol. Chem.* **2009**, *284*, 6446.
(21) Massi, F.; Grey, M. J.; Palmer, A. G., III. *Protein Sci.* **2005**, *14*, 735.
(22) Ban, D.; Funk, M.; Gulich, R.; Egger, D.; Sabo, T. M.; Walter, K. F. A.; Fenwick, R. B.; Giller, K.; Pichierri, F.; de Groot, B. L.; Lange, O. F.; Grubmüller, H.; Salvatella, X.; Wolf, M.; Loidl, A.; Kree, R.; Becker, S.; Lakomek, N.-A.; Lee, D.; Lunkenheimer, P.; Griesinger, C. *Angew. Chem., Int. Ed.* **2011**, *50*, 11437.
(23) Salvi, N.; Ulzega, S.; Ferrage, F.; Bodenhausen, G. *J. Am. Chem. Soc.* **2012**, *134*, 2481.
(24) Tollinger, M.; Sivertsen, A. C.; Meier, B. H.; Ernst, M.; Schanda, P. *J. Am. Chem. Soc.* **2012**, *134*, 14800.
(25) Ban, D.; Gossert, A. D.; Giller, K.; Becker, S.; Griesinger, C.; Lee, D. *J. Magn. Reson.* **2012**, *221*, 1.
(26) Briggman, K. B.; Tolman, J. R. *J. Am. Chem. Soc.* **2003**, *125*, 10164.
(27) Peti, W.; Meiler, J.; Brüschweiler, R.; Griesinger, C. *J. Am. Chem. Soc.* **2002**, *124*, 5822.
(28) Lakomek, N. A.; Walter, K. F. A.; Fares, C.; Lange, O. F.; de Groot, B. L.; Grubmuller, H.; Bruschweiler, R.; Munk, A.; Becker, S.; Meiler, J.; Griesinger, C. *J. Biomol. NMR* **2008**, *41*, 139.
(29) Salmon, L.; Bouvignies, G.; Markwick, P.; Lakomek, N.; Showalter, S.; Li, D. W.; Walter, K.; Griesinger, C.; Bruschweiler, R.; Blackledge, M. *Angew. Chem., Int. Ed.* **2009**, *48*, 4154.
(30) Salmon, L.; Bouvignies, G.; Markwick, P.; Blackledge, M. *Biochemistry* **2011**, *50*, 2735.
(31) Lienin, S. F.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1998**, *120*, 9870.
(32) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. *J. Am. Chem. Soc.* **1995**, *117*, 12562.
(33) Chang, S. L.; Tjandra, N. *J. Magn. Reson.* **2005**, *174*, 43.
(34) Ferrage, F.; Pelupessy, P.; Cowburn, D.; Bodenhausen, G. *J. Am. Chem. Soc.* **2006**, *128*, 11072.
(35) Reese, M.; Turke, M. T.; Tkach, I.; Parigi, G.; Luchinat, C.; Marquardsen, T.; Tavernier, A.; Hofer, P.; Engelke, F.; Griesinger, C.; Bennati, M. *J. Am. Chem. Soc.* **2009**, *131*, 15086.
(36) Morris, G. A.; Freeman, R. *J. Am. Chem. Soc.* **1979**, *101*, 760.
(37) Burum, D. P.; Ernst, R. R. *J. Magn. Reson.* **1980**, *39*, 163.
(38) Pelupessy, P.; Espallargas, G. M.; Bodenhausen, G. *J. Magn. Reson.* **2003**, *161*, 258.
(39) Pelupessy, P.; Ferrage, F.; Bodenhausen, G. *J. Chem. Phys.* **2007**, *126*, 134508.
(40) Ferrage, F.; Cowburn, D.; Ghose, R. *J. Am. Chem. Soc.* **2009**, *131*, 6048.
(41) Ferrage, F.; Reichel, A.; Battacharya, S.; Cowburn, D.; Ghose, R. *J. Magn. Reson.* **2010**, *207*, 294.
(42) Walker, O.; Varadan, R.; Fushman, D. *J. Magn. Reson.* **2004**, *168*, 336.
(43) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836.
(44) Liu, Z.; Zhang, W. P.; Xing, Q.; Ren, X. F.; Liu, M. L.; Tang, C. *Angew. Chem., Int. Ed.* **2012**, *51*, 469.
(45) Ryabov, Y. E.; Fushman, D. *J. Am. Chem. Soc.* **2007**, *129*, 3315.
(46) Kay, L. E.; Nicholson, L. K.; Delaglio, F.; Bax, A.; Torchia, D. A. *J. Magn. Reson.* **1992**, *97*, 359.
(47) Palmer, A. G., III; Skelton, N. J.; Chazin, W. J.; Wright, P. E.; Rance, M. *Mol. Phys.* **1992**, *75*, 699.
(48) Fushman, D.; Cahill, S.; Cowburn, D. *J. Mol. Biol.* **1997**, *266*, 173.
(49) Showalter, S. A.; Bruschweiler, R. *J. Chem. Theory Comput.* **2007**, *3*, 961.
(50) Maragakis, P.; Lindorff-Larsen, K.; Eastwood, M. P.; Dror, R. O.; Klepeis, J. L.; Arkin, I. T.; Jensen, M. O.; Xu, H. F.; Trbovic, N.; Friesner, R. A.; Palmer, A. G.; Shaw, D. E. *J. Phys. Chem. B* **2008**, *112*, 6155.
(51) Markwick, P. R. L.; Bouvignies, G.; Salmon, L.; McCammon, J. A.; Nilges, M.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 16968.
(52) Fenwick, R. B.; Esteban-Martin, S.; Richter, B.; Lee, D.; Walter, K. F. A.; Milovanovic, D.; Becker, S.; Lakomek, N. A.; Griesinger, C.; Salvatella, X. *J. Am. Chem. Soc.* **2011**, *133*, 10336.
(53) Long, D.; Brüschweiler, R. *PLoS Comput. Biol.* **2011**, *7*, e1002035.
(54) Bouvignies, G.; Bernado, P.; Meier, S.; Cho, K.; Grzesiek, S.; Bruschweiler, R.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13885.
(55) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. *Nature* **2007**, *450*, 913.
(56) Kroenke, C. D.; Loria, J. P.; Lee, L. K.; Rance, M.; Palmer, A. G., III. *J. Am. Chem. Soc.* **1998**, *120*, 7905.
(57) Wong, V.; Case, D. A.; Szabo, A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11016.